Charles University in Prague

Faculty of Science

Department of Biochemistry

Biochemistry

Mgr. Tomáš Kučera

Impact of the glycine-rich loop on the function of processing peptidases of the mitochondrial type

Ph.D. Thesis

Supervisor: Ing. Jiří Janata, CSc.

Carried out at: Institute of Microbiology
The Academy of Sciences of The Czech Republic

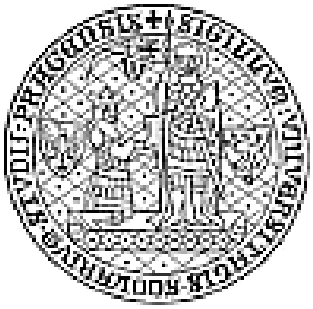
Prague, 2014

Univerzita Karlova v Praze

Přírodovědecká fakulta

Katedra biochemie

Biochemie



Mgr. Tomáš Kučera

Vliv glycinové smyčky na funkci „processing“ peptidas mitochondriálního typu

Disertace

Školitel: Ing. Jiří Janata, CSc.

Vypracováno: Mikrobiologický ústav
Akademie věd České republiky

Praha, 2014

P R O C L A M A T I O N

I declare that I carried out this dissertation independently, and only with the cited sources, literature and other professional sources. This thesis nor its substantial part has been submitted to obtain the same or another academic degree.

In Prague, March 26th 2014

Mgr. Tomáš Kučera

P R E F A C E

I would like to take this opportunity to express my thanks to the following persons who have been supporting me either emotionally, financially or by both during the years of my PhD study. Namely, I would like to thank sincerely to the following persons:

To my parents and my whole family for their support they have been giving me in every moment of my live, whatever I decide to do.

To my supervisor Dr. Jiri Janata for the understanding, generosity, friendliness and freedom. And last but not least for scientific supervision and financial support.

To Dr. Michal Otyepka for advices on molecular dynamics simulation, to Mgr. Petr Pachl for help with protein crystallization, to Dr. Petr Man, Dr. Petr Novák and Mgr. Zdeněk Kukačka for cooperation on mass spectroscopy-related techniques.

To all members of our lab – to Lucka, Eva, Anna, Lenka, Standa, Natálka, Tomáš, Gábina, Abdul and Bojana, for friendliness and for cheerful not-working-only surrounding.

To all my friends, either also PhD students or other friends not affected by the "PhD syndrome". Our more or less similar stories made us even stronger friends.

The list of people to whom I own big thanks is endless. Without your support I wouldn't be writing these lines. Thank you.

A B S T R A C T

The majority of the mitochondrial proteins is synthetized on the cytosolic ribosomes in the form of the protein precursors bearing mitochondrion-targeting signal presequences. Once the protein precursor has reached the mitochondrial matrix the signal presequence is no longer necessary and is cleaved off by heterodimeric mitochondrial processing peptidase (MPP; α/β). Although the crystal structure of MPP is available, the MPP mechanism of function is still matter of discussion.

An all atomic, non-restrained molecular dynamics (MD) simulation in explicit water was used to study in detail the structural features of the highly conserved glycine-rich loop (GRL) of the regulatory α-subunit of the yeast MPP. Wild-type and GRL-deleted MPP structures were studied both in the presence and absence of a substrate in the peptidase active site. Targeted MD simulations were employed to study the mechanism of substrate translocation from the GRL to the peptidase active site. We demonstrate that the natural conformational flexibility of the GRL is crucial for the substrate translocation process from outside the enzyme towards the MPP active site. We show that the α-helical conformation of the substrate is important not only during its initial interaction with MPP (i.e. substrate recognition), but also later, at least during the first third of the substrate translocation trajectory. Further, we show that the substrate remains in contact with the GRL during the whole first half of the translocation trajectory where hydrophobic interactions play a major role. Finally, we conclude that the GRL acts as a precisely balanced structural element, holding the MPP subunits in a partially closed conformation regardless the presence of a substrate in the active site.

Hydrogenosomes are evolutionary related reduced versions of mitochondria that possess MPP-like peptidase – hydrogenosomal processing peptidase (HPP; αβ). We show that HPP is functional as a heterodimer consisting of a regulatory α-subunit and catalytic β-subunit and processes the same set of signal presequences as MPP. On contrary to the MPP, the crystal structure of HPP

has not yet been solved. Beside of the crystallization attempts we employed advanced methods of structural biology to study the structural features of HPP. Specifically, using biological small-angle X-ray scattering and hydrogen-deuterium exchange methods we show that HPP does not undergo conformation changes depending on whether the peptide substrate is or is not present in the peptidase active site. We show that while the wild-type HPP resembles very likely the same quaternary organization as MPP, the structure of the proteolytically inactive E56Q mutant differs from the wild-type one. Chemical cross-linking confirms the low validity of HPP homologous model and suggests that β-HPP is capable of forming homodimers.

With regard to the idea that mitochondria are of the α-proteobacterial origin and to the fact that MPP-like peptidases have been found also in bacteria we worked out an evolutionary scenario for MPP. While MPP and very likely also HPP contain the "full length" GRL, the bacterial MPP-like peptidases contain only the "embryonal" GRL, if any, which is consistent with their biological functions. We conclude that the presence or absence of the "full length" GRL can be considered as evolutionary marker of the physiological function of the given MPP-like peptidase. Furthermore, we hypothesize that the ancestral MPP was attached to mitochondrial inner membrane in the form similar to the *Sphingomonas sp*. heterodimeric peptidase where the prolongation of GRL from the "embryonal" to the "full length" form occurred and, thus, the MPP physiological function of processing peptidase was established. In this context, we also conclude that MPP is an illustrative example of the organelle-driven evolution of the eukaryotic cell.

S O U H R N

Většina mitochondriálních proteinů je syntetizována na cytosolických ribozomech ve formě svých prekurzorů nesoucích signální sekvence, které umožňují jejich transport do mitochondrií. Jakmile proteinový prekurzor dosáhne mitochondriální matrix, signální sekvence již není potřeba a je odštěpena heterodimerní mitochondriální „processing" peptidasou (MPP; α/β). Ačkoli krystalová struktura MPP je známa, mechanismus funkce MPP je stále předmětem diskusí.

Volná molekulárně-dynamická (MD) simulace byla použita k detailnímu studiu strukturních znaků glycinové smyčky (GRL) regulační α-podjednotky kvasinkové MPP. Struktury divoké a mutantní formy MPP s delecí celé glycinové smyčky byly studovány i v přítomnosti substrátu v aktivním místu peptidasy. Cílená MD simulace byla použita ke studiu mechanismu translokace substrátu z GRL do aktivního místa. Prokázali jsme, že přirozená konformační flexibilita GRL je zcela zásadní pro translokaci substrátu z okolního prostředí do aktivního místa peptidasy. Ukazujeme, že α-helikální konformace substrátu je důležitá nejen během jeho prvotního kontaktu s MPP (t.j. rozpoznání substrátu), ale také později, přinejmenším během prví třetiny translokační dráhy substrátu. Dále ukazujeme, že substrát zůstává v kontaktu s GRL během celé první třetiny translokace, během níž hydrofóbní kontakty hrají zásadní roli. Nakonec shrnujeme, že GRL je jemně vyvážený struktuní prvek, který drží podjednotky MPP v částečně uzavřené konformaci, bez ohledu na přítomnost substrátu v aktivním místě peptidasy.

Hydrogenosomy jsou evolučně příbuzné redukované formy mitochondrií, které obsahují peptidasu typu MPP – hydrogenosomální „processing" peptidasu (HPP; α/β). Ukazujeme, že HPP je funkční jako heterodimer skládající se z regulační α- and katalytické β-podjednotky, a že zpracovává stejnou sadu substrátů jako MPP. Na rozdíl od MPP, krystalová struktura HPP nebyla dosud vyřešena. Kromě krystalizačních experimentů jsme využili pokročilé metody strukturní biologie ke studiu strukturních rysů HPP. Konkrétně, užitím metody biologického rozptylu rentgenových paprsků a metody výměny vodíkových

atomů za deuterium ukazujeme, že HPP nevykazuje konformační změny v závislosti na přítomnosti substrátu v aktivním místě. Ukazujeme, že zatímco divoká forma HPP má velmi pravděpodobně stejnou kvartérní organizaci jako MPP, tak struktura proteolyticky neaktivní formy HPP se liší od divoké.

S ohledem na představu, že mitochondrie jsou α-proteobakteriílního původu, a s ohledem na fakt, že peptidasy typu MPP byly nalezeny také v bakteriích, jsme vypracovali evoluční scénář pro MPP. Zatímco MPP a, velmi pravděpodobně také, HPP obsahují „plnou" GRL, tak bakteriální peptidasy typu MPP obsahují pouze „embryonální" GRL, pokud vůbec, což souhlasí s jejich biologickými funkcemi. Shrnujeme, že přítomnost „plné" GRL může být považována za evoluční značku fyziologické funkce dané peptidasy typu MPP. Dále shrnujeme, že předchůdce MPP byl připojen ke vnitřní membráně mitochondrie ve formě podobné heterodimerní peptidase ze *Sphingomonas sp*., kde proběhlo prodloužení GRL ze své „embryonální" do „plné" podoby, a tedy, kde se ustanovila funkce „processing" peptidasy. V tomto smyslu dále shrnujeme, že MPP je ukázkový příklad evoluce eukaryotické buňky řízené organelou.

CONTENT

1. INTRODUCTION 1

2. LITERATURE REVIEW 4

2.1 EVOLUTION OF THE EUKARYOTIC CELL 4

2.1.1 Mitochondrion and its origin 5

2.1.1.1 Conventional endosymbiotic theory 5

2.1.1.2 Hydrogen hypothesis 6

2.1.1.3 Ox-tox hypothesis 8

2.1.1.4 Syntrophy hypothesis 9

2.1.2 Hydrogenosome and its relation to mitochondrion 10

2.1.3 Host cell and its origin 13

2.2 EVOLUTION OF MITOCHONDRIAL PROTEIN-IMPORT MACHINERY 17

2.2.1 The origin of targeting peptides 19

2.2.2 Mitochondrial targeting peptides 20

2.2.3 Transport of mitochondrial proteins 21

2.3 M16 FAMILY PEPTIDASES 25

2.3.1 Mitochondrial processing peptidase 27

2.3.1.1 Structural features 27

2.3.1.2 Substrate recognition 28

2.3.1.3 Evolutionary context 29

2.3.2 Hydrogenosomal processing peptidase 31

2.3.2.1 Evolutionary context 32

2.3.3 Bacterial M16B peptidases 33

2.3.3.1 Rickettsia prowazekii peptidase 34

2.3.3.2 Bacillus halodurans peptidase 37

2.3.3.3 Thermus thermophilus peptidase 38

2.3.3.4 Sphingomonas sp. peptidase 40

3. AIMS 42

3.1 THE ROLE OF GRL IN MPP STRUCTURE AND FUNCTION 42

3.2 STRUCTURAL FEATURES OF HPP 42

3.3 EVOLUTIONARY PATH OF MPP AND ITS GRL 43

4. MATERIAL AND METHODS 44

4.1 MATERIAL 44

4.1.1 Cloning vectors and constructs 44

4.1.2 Bacterial strains 45

4.1.3 Material *46*
4.2 DNA MANIPULATIONS 46
4.2.1 General DNA techniques *46*
4.2.2 Site-directed mutagenesis *47*
4.2.3 pET Duet expression system *47*
4.3 PROTEIN EXPRESSION AND PURIFICATION 47
4.3.1 Protein expression *47*
4.3.2 Metal affinity purification *48*
4.3.3 Protein concentration and storage *49*
4.3.4 Gel chromatography *50*
4.3.5 Activity assay and protein quantification *51*
4.3.6 Western Blot *51*
4.3.7 Differential scanning fluorimetry *51*
4.4 CONSTRUCTION OF THE HOMOLOGY MODEL OF HPP 53
4.5 PROTEIN CRYSTALLOGRAPHY 53
4.5.1 Dynamic light scattering *53*
4.5.2 Screening for crystallization conditions *54*
4.6 BIOLOGICAL SMALL-ANGLE X-RAY SCATTERING (BIO-SAXS) 55
4.6.1 Samples preparation *55*
4.6.2 Experimental setup *55*
4.6.3 Data processing *56*
4.7 HYDROGEN-DEUTERIUM (H/D) EXCHANGE 56
4.7.1 H/D exchange experiment *56*
4.7.2 Digestion and HPLC separation *57*
4.7.3 Mass spectrometry and data analysis *57*
4.7.4 MALDI-TOF substrate cleavage assay *58*
4.8 PROTEIN CROSS-LINKING TECHNIQUES 58
4.8.1 Cross-linking experiment *58*
4.8.2 Sample digestion and LC-MS analysis *59*
4.8.3 Data analysis *59*
4.9 MOLECULAR DYNAMICS SIMULATIONS 60
4.9.1 Models for substrate translocation study *60*
4.9.1.1 The GRL-bound structure 60
4.9.1.2 The AS-bound structure 60
4.9.2 Models for the study of the structural role of GRL *61*

4.9.3 Targeted MD simulations *62*
4.9.4 Non-restrained MD simulations *62*
4.9.5 Data analysis *64*
5. RESULTS **65**
5.1 HPP EXPRESSION AND PURIFICATION 65
5.1.1 Co-expression and pET Duet expression *65*
5.1.2 Metal affinity purification of WT and E56Q HPP dimer *65*
5.1.3 Purification of the mutant forms of a-HPP subunit *66*
5.1.4 Gel chromatography *68*
5.1.5 Storage conditions - DSF and DLS *69*
5.2 CRYSTALLIZATION EXPERIMENTS 71
5.3 SMALL-ANGLE X-RAY SCATTERING 73
5.3.1 Hydrogenosomal processing peptidase *73*
5.3.2 Mitochondrial processing peptidase *75*
5.4 HYDROGEN-DEUTERIUM EXCHANGE 77
5.4.1 Peptide mapping *78*
5.4.2 MALDI-TOF substrate cleavage assay *79*
5.4.3 Kinetics of H/D exchange for HPP *82*
5.5 CHEMICAL CROSS-LINKING OF HPP 86
5.6 MOLECULAR DYNAMICS SIMULATIONS OF MPP 89
5.6.1 Substrate translocation from GRL to MPP active site *89*
5.6.2 Structural role of GRL for the MPP dimer stability *94*
6. DISCUSSION **99**
6.1 MITOCHONDRIAL PROCESSING PEPTIDASE 99
6.1.1 GRL as an active element during substrate translocation *99*
6.1.2 GRL keeps MPP dimer in a partially closed conformation *102*
6.2 HYDROGENOSOMAL PROCESSING PEPTIDASE 104
6.2.1 Purification procedures *105*
6.2.2 Structural features of HPP *106*
6.3 EVOLUTION OF MPP - DAWN AND FALL OF GRL 110
7. CONCLUSIONS **119**
7.1 THE ROLE OF GRL IN MPP STRUCTURE 119
7.2 STRUCTURAL FEATURES OF HPP 120
7.3 EVOLUTION OF THE GRL OF MPP 121
8. PUBLICATION LIST **123**

8.1 PUBLICATIONS WITH IF 123
8.2 PUBLICATIONS WITHOUT IF 123
9. SUPPLEMNTARY MATERIAL 124
9.1 MULTIMEDIA 124
9.1.1 Substrate translocation from GRL to MPP active site 124
10. REFERENCES 125

ABBREVATIONS LIST

A

ADP - Adenosine Diphosphate.
AK – Adenylate-kinase.
ATP - Adenosine Triphosphate.

B

BHP – Bacillus halodurans Peptidase.

D

DDM – *n*-dodecylmaltoside.

E

EDTA – Ethylenediaminetetraacetic acid.
ER - Endoplasmatic Reticulum.
ESP - Eukaryotic Signature Protein.

G

GA - Golgi Aparatus.
GPP – Giardia Processing Peptidase.
GRL - Glycine-rich Loop.
GRLP – Glycine-rich-loop-like Protein.

H

HPP – Hydrogenosomal Processing Peptidase.

I

IM – Inner Membrane.
IMP – Inner Membrane Peptidase.
IMS – Intermembrane Space.

L

LECA - Last Eukaryotic Common Ancestor.
LUCA - Last Universal Common Ancestor.

M

MIP – Mitochondrial Intermediate Peptidase.
MPD – 2-methyl-2,4-pentanediol.
MPP – Mitochondrial Processing Peptidase.

N

NADH - Nikotinamid Adenin Dinukleotid.

O

OM – Outer Membrane.

P

PAM – Presequence-associated Motor.
PFO – Pyruvat-Ferredoxin Oxidoreductase.
PhAT – Phagocytosing Archaeon Theory.
PreP – Stromal-targeting Peptide Peptidase.

R

RPP – Rickettsia Processing Peptidase.

S

SPH – Sphingomonas sp. Peptidase.
SPP – Stromal Processing Peptidase.

T

TIM – Translocase of Inner membrane.
TOM – Translocase of Outer Membrane.
TTHA – Thermus thermophilus Peptidase.

1. INTRODUCTION

From a scientific view, the life is a characteristic that distinguishes objects that have signaling and self-sustaining processes from those that do not, either because such functions have ceased, or because they lack such functions and are classified as inanimate. Any contiguous living system is called an organism. Organisms undergo metabolism, maintain homeostasis, possess a capacity to grow, respond to stimuli, reproduce and, through natural selection, adapt to their environment in successive generations. More complex living organisms can communicate through various means. Although a diverse array of living organisms can be found in the biosphere of Earth, the common properties of all organisms are a carbon- and water-based cellular form with complex organization and heritable genetic information.

Scientific evidence suggests that life began on Earth approximately 3.5 – 4.2 billion years ago [1, 2]. The mechanism by which life emerged on Earth is unknown although many hypotheses have been formulated. Since then, life has *evolved* into a wide variety of forms, which biologists have hierarchically classified into six groups of taxa - plants, animals, fungi, protists, archaea and bacteria [3]. Evolution is the cornerstone of modern biology. However, the word evolution has a variety of meanings. The fact that all organisms are linked via descent to a common ancestor is often called evolution. The theory of how the first living organisms appeared (i.e. abiogenesis) is also often called evolution. And frequently, the term evolution is often used when talking indeed about natural selection -- one of the many mechanisms of evolution (such as sexual selection or genetic drift).

All organisms on Earth are built up from one of the two primary types of cells – prokaryotic and eukaryotic. Prokaryotes lack a nucleus and other membrane-bound organelles, although they have circular DNA and ribosomes. *Bacteria* and *Archaea* are two domains of prokaryotes. The other primary type of cells are the eukaryotes, which have distinct nuclei bounded by a nuclear

membrane and membrane-bound organelles, including mitochondria, chloroplasts, lysosomes and rough and smooth endoplasmic reticulum. In addition, other mitochondria-like organelles such as hydrogenosomes and mitosomes have also been addressed to eukaryotic cells. All species of large complex organisms are eukaryotes, including animals, plants and fungi, though most species of eukaryote are protist microorganisms.

The conventional model is that eukaryotes evolved from prokaryotes, with the mitochondria and chloroplasts forming through endosymbiosis between bacteria and the progenitor eukaryotic cell. Anyway, this rather historical view has been challenged by many derived hypotheses. Since mitochondrion plays a central role in all evolutionary hypotheses, the first part of this thesis deals with the origin of mitochondria and other mitochondria-like organelles.

Mitochondrion evolution is in turn tightly connected with evolution of mitochondrial protein import machinery. Thus, the second part of the thesis deals with the origin of mitochondria-targeting sequences and theirs structural features. Subsequently, attention is paid to the questions how these signal sequences are recognized by the import protein complexes and how are the mitochondrial proteins transported across the mitochondrial membranes.

The last part of the introduction focuses on the processes subsequent to the transport through mitochondrial membranes. In other words, this is the moment where mitochondrial processing peptidase (MPP) comes into play. The crucial role of MPP is cleaving off the signal presequence and thus allowing the transported proteins to fold into native conformation. To fulfill its function, MPP features a structural element called glycine-rich loop (GRL) which recognizes signal presequences prior to the subsequent proteolysis. However, GRL was found also in bacterial and hydrogenosomal MPP-like peptidases, though often in "embryonal", reduced or modified form. Thus, GRL may provide clues to the evolutionary origin and fate of not only MPP but also mitochondrion and whole eukaryotic cell.

To sum it up, the ultimate goal of presented work is an attempt to trace the evolution of glycine-rich loop of MPP and MPP-like peptidases together with their physiological functions in the context of the eukaryotic cell evolution.

2. LITERATURE REVIEW

2.1 Evolution of the eukaryotic cell

It was generally proposed that eukaryotes evolved from prokaryotes in the pre-genomic era. The logic of this thinking is simple: prokaryotes are simple and do not possess many features that eukaryotes do, i.e., the nucleus, endoplasmic reticulum (ER), Golgi apparatus (GA), mitochondria, a complex cytoskeletal network, and prokaryotes do not have sexual reproduction. However, the rRNA sequence comparisons revealed the existence of two different kinds of prokaryotes (eubacteria and archaebacteria) [4]. Thus, the historical view has been challenged and later the three domain classification of life was established: Bacteria (eubacteria), Archaea (archaebacteria), and Eukarya (eukaryotes) [5]. Although this three domain classification (**Figure 1**) is still accepted today, the evolutionary origin of the eukaryotic cell represents an enigmatic, yet largely incomplete, puzzle.

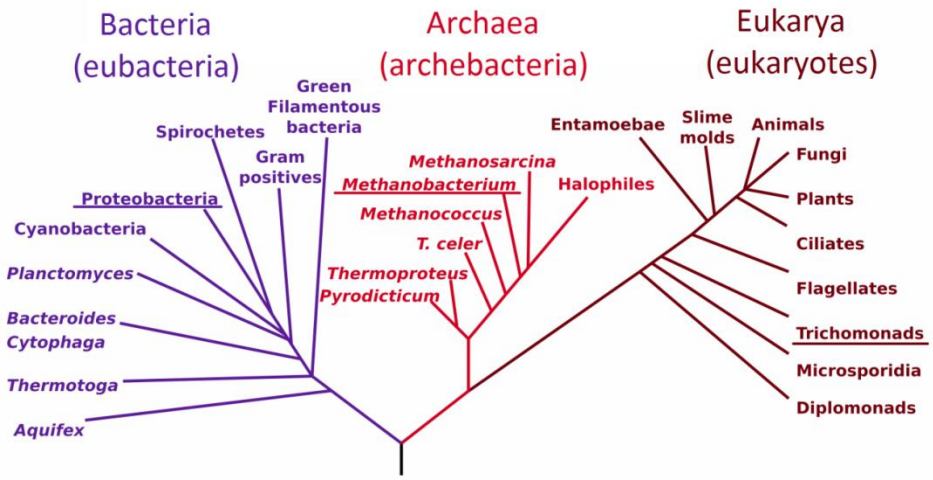


Figure 1. Phylogenetic tree of life. Underlined orders are mentioned further in the text.

The question of the evolution of the eukaryotic cell may be viewed traditionally from two perspectives – from the perspective of the mitochondrion and from the perspective of the host cell. Although both perspectives is not possible to separate completely, from the historical and systematical reasons they will be described in two separate chapters.

2.1.1 Mitochondrion and its origin

The acquisition of mitochondria was an important event in the evolution of the eukaryotic cell, supplying it with compartmentalized bioenergetic and biosynthetic factories. Advances in geochemistry, molecular phylogeny, and cell biology have offered insight into complex of molecular events that drove the evolution of endosymbiont into contemporary organelle. In losing their autonomy, endosymbiont lost the bulk of its genomes, necessitating the evolution of mechanisms for organelle biogenesis and metabolite exchange. In the process, symbiont acquired many host-derived properties, lost much of its identity, and was transformed into what is today known as mitochondria, the essential parts of the eukaryotic cells.

Several hypotheses have been postulated in the course of time with the goal to decipher the mitochondrion origin, leading eventually to the now widely accepted idea that the mitochondrion is of α-proteobacterial origin. The endosymbiotic theory was postulated as the first one and in its original form is today called "conventional". Subsequently, several newer hypotheses have been proposed that extend the conventional endosymbiotic theory and thus bring new interesting insights into the mystery of mitochondrion origin. Although these new theories also belong principally to the parental conventional endosymbiotic theory, they will be described here as independent standalone hypotheses, since they differ in some important aspects.

2.1.1.1 Conventional endosymbiotic theory

The conventional endosymbiotic theory has a special position since this evolutionary scenario was the subject of serious debates already in the 19th century and more recently was revived by Lynn Margulis [6]. According to this theory, the symbiont obtained metabolizable substrates and physical protection in exchange for respiration-derived ATP. The later analysis of mitochondrial genes and their genomic organization and distribution revealed that mitochondrial genes are of

bacterial origin and are derived most likely form an α-proteobacterium-like ancestor [7, 8]. Indeed, the analysis of the first completed genome sequence of α-proteobacterium *Rickettsia prowazekii* has provided additional support for the idea that mitochondria are derived from a free-living α-proteobacterium that established an endocellular abode in a primitive ancestor to the modern eukaryotic cell [9, 10]. In time, this bacterial endosymbiont was reduced to a highly dependent organelle with the vast majority of its genes either lost or transferred to the eukaryotic nucleus [11].

Informative examples of mitochondrial proteins of non-bacterial origin are the ATP/ADP translocases, which export ATP in exchange for ADP across membranes. Surprisingly, there are no similarities between the rickettsial and mitochondrial transporters. This suggests that the ATP/ADP transport function has independently arisen twice. Moreover, there are no logical reasons to expect a free living ancestral bacterium to either export or import ATP to or from its environment. Thus, this suggests that the ancestral α-proteobacteria was not capable of ATP transport at the time of endosymbiotic event [9, 12]. Finally, all the ATP/ADP translocases of mitochondria are encoded in cell nucleus. These facts point to the conclusion that although ATP production in mitochondria originated from an α-proteobacteria ancestor, the transmembrane ATP transport appears to have originated in the eukaryotic genome subsequent to the divergence of mitochondria and α-proteobacteria.

However, if ATP was not exchanged during the initial endosymbiotic event, what supported the original symbiotic relationship and what was the driving force of the process as a whole?

2.1.1.2 Hydrogen hypothesis

The hydrogen hypothesis postulates that hydrogen rather than ATP was the molecular bond that supported the endosymbiotic association. According to this scenario, the host was an anaerobic, strictly hydrogen-dependent and strictly

autotrophic archaebacterium while the symbiont was an eubacterium capable of respiration, but generating molecular hydrogen as a waste product of anaerobic heterotrophic metabolism (**Figure 2**) [13]. This scenario suggests that the two organisms first met in an environment that was rich in hydrogen and carbon dioxide, and one that could support a facultative anaerobe. An association between the two organisms was established subsequently based on the production of hydrogen and carbon dioxide by the symbiont. Thus, the archaeon became more and more dependent on its symbiont, which eventually was completely engulfed. The theory further suggests that the mitochondrion ancestor carried all enzymes found in current mitochondria and hydrogenosomes and their separate distribution in contemporary organisms is explained by differential gene losses. In other words, the loss of anaerobic metabolism in aerobic environment led to the origin of mitochondria, whilst the loss of the respiration in anaerobic environment directed the endosymbiont evolution to hydrogenosome.

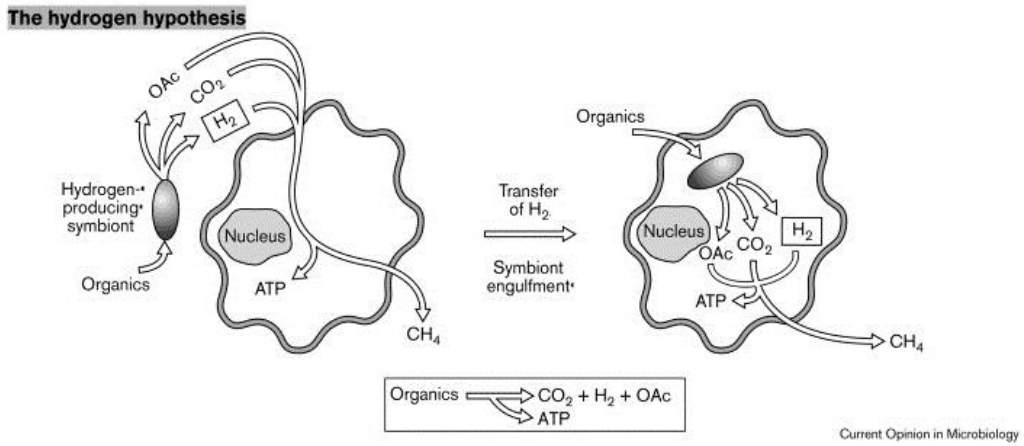


Figure 2. A schematic view of the hydrogen hypothesis for the origin of aerobic respiration in eukaryotes [13]. It is suggested that the initial symbiobitic relation leading to mitochondria was supported by the transfer of hydrogen from a hydrogen-producing bacterium to an anaerobic, hydrogen-dependent host. Here, mitochondria are thought to have been acquired simultaneously with the origin of the eukaryotic lineage. OAc – acetyl-oxy group. Figure taken from [14].

In effect there are two important differences between the conventional and hydrogen hypothesis:

- The aerobic versus the facultative anaerobic nature of the symbiont.

- The utilization of ATP as opposed to hydrogen to control the initial symbiosis.

However, recently it was summarized that molecular phylogenetic data have "confirmed the simplest version of the endosymbiosis hypothesis", which addresses the origin of aerobic ATP-producing pathways in mitochondria [14]. Furthermore, it was concluded that the hydrogen hypothesis for the origin of mitochondria, which addresses the origins of anaerobic ATP producing pathways in hydrogenosomes, does not receive support from molecular data.

Thus, although the hydrogen theory described above is supported by a considerable amount of evidence, cannot be universally accepted as the explanation of the common origin of mitochondrion and hydrogenosome.

2.1.1.3 Ox-tox hypothesis

Ox-tox hypothesis for the origin of mitochondria is based on a two-phase selection for aerobic respiration by an α-proteobacterial symbiont (**Figure 3**) [15]. Here, the initial function of the symbiont is the detoxification and the later function is the provision of ATP to the host cell. In this scenario, the precipitous rise in atmospheric oxygen about 200 million year ago is recognized as an environmental disaster for anaerobic organisms. These could survive by withdrawing to dwindling anaerobic environments, or, could adapt to the presence of oxygen by exploiting the capacity of facultative aerobic bacteria to locally detoxify the niche by consuming oxygen [16].

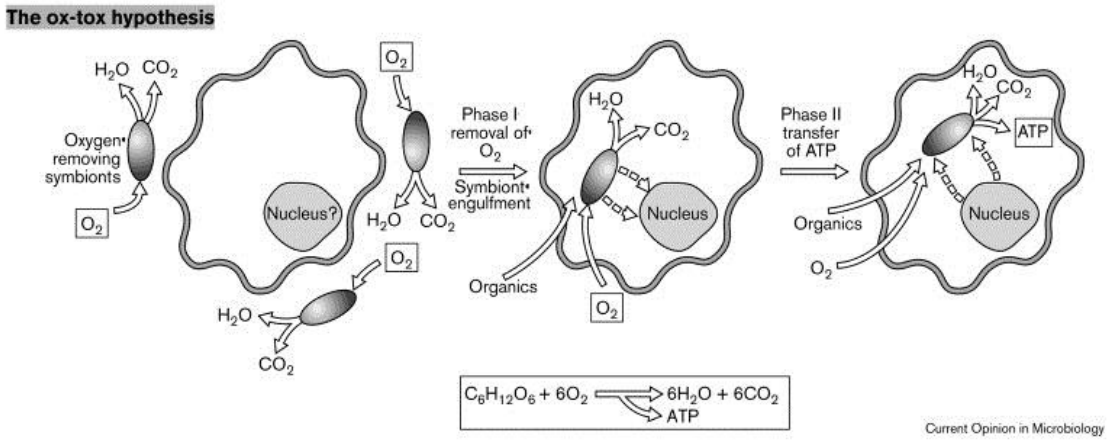


Figure 3. A schematic view of the ox-tox hypothesis for the origin of aerobic respiration in eukaryotes [15]. Here, the acquisition of mitochondria is based on the symbiosis established by a bacterium in two phases. In phase I, the symbiont detoxifies the host cytoplasm by consuming oxygen. During phase II, the transport of ATP from the mitochondrion to the host cell is implemented by the acquisition of appropriate proteins encoded by the host nuclear genome. The dashed arrows represent the transfer of bioenergetic and information genes from the proto-mitochondrion to the nuclear genome (phase I) and the evolution of novel genes in the nuclear genome for mitochondrial functions (phase II). Figure taken from [14].

Such a commensal association could become more intimate if the aerobe took residence within its host. At this stage the symbiont would function to detoxify the intracellular environment of the anaerobic host. In the second phase, the integration of their metabolism would proceed by the symbiont's acquisition of the host transport and control functions, such as the acquisition of an ATP/ADP translocase complexes encoded by the host genome. Thus, genes such as those encoding the mitochondrial ATP/ADP translocases, and other proteins involved in transport and cellular communication seem to signal the transition from an oxygen-scavenging to ATP producing function for the evolving mitochondrion.

2.1.1.4 Syntrophy hypothesis

The syntrophy hypothesis is also based on interspecies hydrogen transfer but takes into account different organisms (**Figure 4**) [17, 18]. The hydrogen and ox-tox hypotheses also differ from the syntrophy hypothesis in the number of anticipated endosymbiotic events: the hydrogen and ox-tox hypotheses propose a

single endosymbiotic event, whereas the syntrophy hypothesis postulates two separate events. The first symbionts are suggested to be sulfate-reducing myxobacteria (Bacteria) that established a syntrophic consortium with methanogens (δ-proteobacteria) based on hydrogen transfer. The symbiont became the nucleus in the resulting, primitively amitochondriate eukaryote. In second step, this cell then acquired an anaerobic methanotrophic α-proteobacterium that was feeding on the methane produced by the methanogenic consortium and eventually became the mitochondrion.

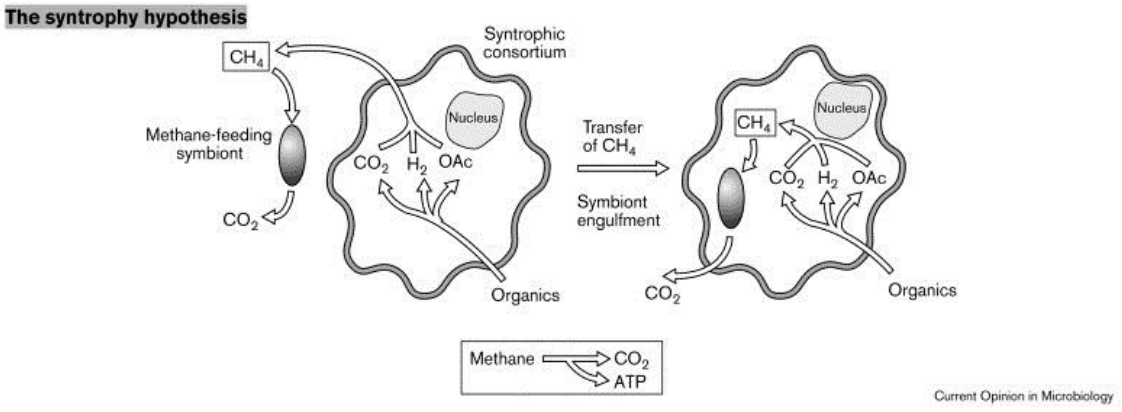


Figure 4. A schematic view of the syntrophy hypothesis for the origin of aerobic respiration in eukaryotes [17]. In this model, the host is a syntrophic consortium consisting of hydrogen-producing bacteria and anaerobic, hydrogen-dependent methanotrophs. The symbiont that ultimately becomes the mitochondrion is initially an anaerobic methanotroph that depends on the methane produced by the syntrophic consortium. Here, it is proposed that the mitochondrion originated subsequent to the establishment of the eukaryotic cell. Figure taken from [14].

2.1.2 Hydrogenosome and its relation to mitochondrion

Several eukaryotic organisms do not have mitochondria but possess (a) double-membraned organelles called hydrogenosomes, which (b) produce ATP fermentatively. Other typical traits of hydrogenosomes are (c) the production of molecular hydrogen, (d) the biosynthesis of FeS centers and (e) the absence of genome [19]. Reconstructing the evolutionary origin of hydrogenosomes has been hampered by the unavailability of organellar DNA (although exceptions exist [20]), which was central in reconstituting mitochondrial origin. Thus, the defining

the mitochondrion-hydrogenosome relationship is circumstantial and must be proteome-based.

Unlike typical aerobic mitochondria, which use pyruvate dehydrogenase for pyruvate oxidation, *Trichomonas* hydrogenosomes decarboxylate pyruvate with pyruvate:ferredoxin oxidoreductase (PFO) which transfers electrons to a [Fe]-hydrogenase, ultimately producing ATP, H_2 and CO_2 [21]. PFO and hydrogenase, for which there are no counterparts in mitochondria, are typically found only in anaerobic bacteria, and the origin of the eukaryotic homologs is unknown, although it appears that eukaryotic PFO has a single origin [22] (however, not traceable down to an α-proteobacterial donor). However, phylogenetic analyses of a few protein-coding genes have suggested a common ancestry for hydrogenosomes and mitochondria, as do similarities in organelle biogenesis [19].

As it has been already outlined, the publication of the genome sequence from the parasitic α-proteobacterium *R. prowazekii* has revealed a complete set of tricarboxylic cycle and respiratory-chain complex enzymes that are phylogenetically related to their mitochondrial counterparts [9, 10]. Thus, it would seem that *R. prowazekii* and the mitochondrion had a common ancestor. A body of evidence also indicates that hydrogenosomes and mitochondria share a common ancestor (i.e. hydrogenosomes are anaerobic forms of mitochondria) [23-25]. However, typical hydrogenosomal enzymes, such as hydrogenase and PFO, are not found in *R. prowazekii* [9], suggesting either that (a) these have been lost from the highly reduced *R. prowazekii* genome, or alternatively, that (b) two different endosymbiotic events were involved in the establishment of mitochondria and hydrogenosomes.

This issue invites discussion from an alternative perspective. The conventional endosymbiotic theory and its derived hypotheses account for the origin of aerobic mitochondria only – not for the origin of anaerobic mitochondria nor hydrogenosomes. In fact, all endosymbiotic models for the origin of mitochondria focus on the derivation of a narrow and specific subset of

mitochondrial diversity – namely typical, textbook-like, aerobic mitochondria such as those found in cells of human liver.

Such organelles utilize pyruvate dehydrogenase for oxidative decarboxylation, citric acid cycle to generate NADH, that is fed into the ATP-producing respiratory chain with O_2 serving as a terminal acceptor. The same biochemistry was found in the obligate aerobe *R. prowazekii*. However, *Rickettsia* is highly reduced and specialized α-proteobacterium and many free-living α-proteobacteria possess a greater spectrum of biochemical diversity than *Rickettsia*. Similarly, human-liver-type mitochondria are highly specialized organelles. Among eukaryotes that inhabit anaerobic environments and among those that have anaerobic stages in their life cycle, there is a wealth of biochemical diversity in mitochondrial energy metabolism that formulations of the conventional endosymbiotic theory neither account for nor address, arguably because they are designed to explain the origin of an oxygen-consuming organelle.

Novel and intriguing symbiotic models are emerging to account for the differences between prokaryotes and eukaryotes at the level of cellular organization and genome complexity [16, 26]. En example of such a new theory is the syntrophy hypothesis [17, 18] mentioned already in the section dedicated to mitochondrion and its origin. Although these theories have distinctive virtues, they do not directly account for the diversity and compartmentation of ATP-producing pathways among contemporary anaerobic protists.

Thus, how to account for the origin of anaerobic organelles? There are basically two ways to explain the origin of anaerobic biochemistry in mitochondrion and hydrogenosomes. Under one alternative, the ancestral mitochondrion is viewed as oxygen-respiring organelle in adherence to conventional endosymbiotic theory, and the genes for enzymes specific to ATP synthesis in anaerobic mitochondria and hydrogenosomes (such as PFO are viewed as acquisitions involving independent lateral gene transfer events in different eukaryotic lineages [14]. Under a different alternative, the common

ancestor of mitochondria and hydrogenosome is viewed as a facultatively anaerobic α-proteobacterium that was able to satisfy its ATP needs with and without the help of oxygen, whereby the imprint of this facultatively anaerobic past is preserved in the spectrum of organelle diversity that is observed among protists today [13]. Unfortunately, the available data today are too limited to distinguish between these two scenarios for the origin of anaerobic organelles. The study of eukaryotes that do not depend upon oxygen may provide the incisive clues.

2.1.3 Host cell and its origin

Today there is little doubt that mitochondria monophyletically arose from within α-proteobacteria because comparative genome data permit no other interpretation [9, 11, 27]. Seen from the perspective of the ancestral α-proteobacterium, the mitochondrion has evolved both by extensive loss of ancient coding sequences as well as by a more selective acquisition of eukaryotic proteins. However, all endosymbiotic theories has fared much better when it comes to explaining the origins of organelles that it has when it comes to explaining the origin of their host [28, 29].

The last universal common ancestor (LUCA) of all organisms is currently a matter of wide discussion. It might have been a bacterium [30], an archaeon [31], an eukaryote [32] or something completely different lacking tight linkage between genotype and phenotype [33], self-replicating pre-cells unable to control the exchange of genetic information or even proto-eukaryotes possessing at least some eukaryotic features [34]. Eukaryotes are assumed to have arisen by genetic or even cellular intermixing of two or more members of different domain. This might be true since interdomain as well as intradomain lateral gene transfer is a fact and all eukaryotes possess mitochondria or at least their remnants if the forms of hydrogenosomes or mitosomes [35]. Beyond this, the ideas are blurry.

Despite the large number of mutually incompatible evolutionary hypotheses for the origin of eukaryotes, some consensus has been reached about the following points:

- The last eukaryotic common ancestor (LECA) contained the mitochondrial progenitor derived from endosymbiosis with an α-proteobacterium (as discussed in detail in chapter 2.1.1).
- Eukaryotic genomes have a chimeric nature: genes for information storage and processing are Archaea-related, and genes for metabolic or "operational" processes are mostly bacterial in nature (but not necessarily derived from the mitochondrial progenitor).
- A significant fraction of the eukaryotic genes encode proteins that are restricted to eukaryotes, the so-called ESPss (eukaryotic signature proteins).
- Components of endomembrane system such as ER and GA have a non-endosymbiotic origin.

In addition, following major questions are of interest:

- It is still unclear whether mitochondrion ancestors (thus α-proteobacteria) were ancestrally parasites, obligate symbionts or if they were eaten by host via phagocytosis.
- What was the nature of LUCA, the host that took up the α-proteobacterium?
- When did eukaryotic complexity arise, before ("mitochondria-late") or after ("mitochondria-early") the endosymbiotic event that led to the establishment of the mitochondrion?

Thus, based on the last question the hypotheses for the origin of eukaryotes may be principally classified into two groups. Mitochondria-late models, of which the Archezoa model (in other words, "amitochondriate

eukaryote model") [36] is main protagonist, argue that eukaryotes gradually evolved from a lineage devoid of mitochondria and eventually acquired mitochondria (**Figure 5-A** and **B**). Hence, such scenarios are more conservatively consistent with the historical view and are compatible with the classical three-pronged classification of the domains of life.

In contrast, mitochondria-early models suggest that eukaryotes are direct descendants of neither *Archea* nor *Bacteria* nor their associations or chimeras and argue that the eukaryotic lineage emerged from a symbiosis between an archaeon and a bacterium (the mitochondrial ancestor). Subsequently, this symbiogenesis and/or fusion subsequently triggered the evolution of typical eukaryotic features [13, 37] (**Figure 5-C** and **D**). Thus, this model is more liberal, suggesting that eukaryotes are not simple descendants of prokaryotes, as we know them by definition.

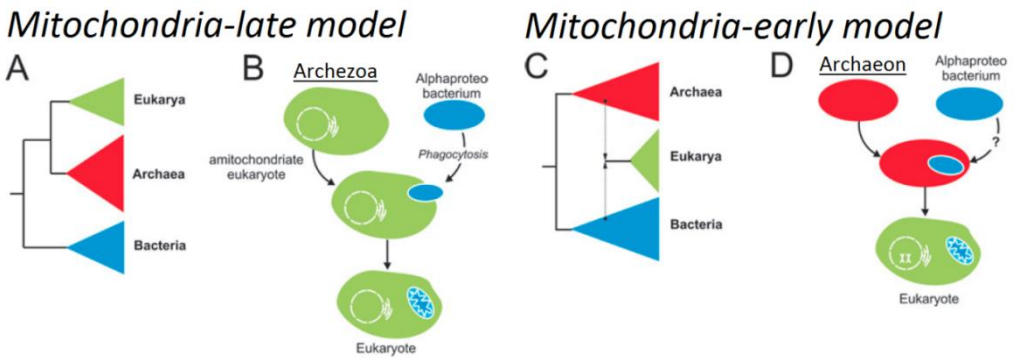


Figure 5. The evolutionary relationships between the three domains of life. Panel A shows schematic phylogenetic tree that displays the classical "three domains" view of life. Panel B shows the scenario in which α-proteobacterial progenitor of mitochondria was acquired relatively late in the eukaryotic evolution. Panel C shows schematic phylogenetic tree displaying a fusion-like origin for the eukaryotic domain of life. Panel D shows the fusion of archaeon with α-proteobacterial progenitor of mitochondria via a hitherto unknown endosymbiotic interaction. Figure taken from [38] and adapted.

Although both models provide a fairly complete explanation, in both cases the cellular intermediates that would support either scenario are unknown. After a brief period of popularity, support for the mitochondria-late scenarios has been decreasing due to the discovery that all previously presumed amitochondriate

eukaryotes contain or were shown to have once contained a mitochondria-derived organelles (hydrogenosomes and mitosomes) [35]. On contrary, recent phylogenetic studies lend support to symbiotic or fusion like scenarios (i.e. mitochondria-early models) at the base of eukaryotic origin.

Specifically, LUCA was most likely an entity bounded by two-membranes of heterochiral lipid composition [39], and possessed ribosomes, membrane secretion and insertion apparatus and ATPase in the inner membrane [40]. LUCA may have had the ability to engage in random fusions and fissions. The lateral gene transfer was probably much more extensive in the era of LUCA then in extant cells and LUCA might have been a self-replicating progenote unable to limit the frequent horizontal exchange of its genetic information [41]. The *Bacteria* domain had most likely arisen first, and the last common ancestor or *Archaea* and *Eukarya* was probably still at the progenote stage, retaining many features of LUCA. While the ancestors of the *Archaea* and *Bacteria* domains probably arose convergently by fixation of enzymes for stereospecific lipid synthesis, emergence of a fusion-prohibiting cell wall, loss of many RNA-based systems and compaction of genomes [41], LECA most likely retained many features tracing back to LUCA. The α-proteobacterial ancestors of mitochondria were probably parasites of pre-karyote intermembrane space.

However, the parasitic nature of mitochondrion progenitor has been recently challenged by "phagocytosing archaeon theory" (PhAT) [38]. The starting point of this hypothesis is an ancestral archaeal lineage belonging to the "TACK superphylum" (comprising *Thaumarchaeota*, *Aigarchaeota*, *Crenarchaeota* and *Korarchaeota*) [42] that is supposed to have contained the full collection of ESPs that have currently been identified in *Archaea*, such as actin [43], tubulin [44], ubiquitin protein modifier machinery [45] and several components of transcription and translation complexes. Subsequently, this ancient TACK lineage lost its proteinacious cell wall, allowing for the evolution of a more flexible actin-based cytoskeleton, such is observed in plasma-like organisms. Next, the cytoskeleton

matured into primitive phagocytosis machinery. As a result, rather than invoking a single acquisition event, the PhAT theory proposes that a significant part of bacterial genes present in eukaryotic genomes originate from phagocytotic ingestions of prokaryotes.

2.2 Evolution of mitochondrial protein-import machinery

The most critical steps in creating mitochondria (i.e. the transition from autonomous endosymbiont to organelle) were the establishment of protein import machinery in the membrane of what was a bacterial symbiont and the linked process of genome reduction. According to the endosymbiotic theories, the genes of bacterial symbiont were either lost [46] or transferred in a gradual process into the host chromosomes.

As it has been already mentioned, it remains unclear what was the metabolic nature of the original symbiosis – whether it was true symbiosis, parasitism or phagocytosis. Therefore, it is difficult to address what factors might have driven the ancient bacterial symbiont to surrender his genome. It remains equally possible that installing protein import machinery enabled the productive transfer of endosymbiont genes to the emerging nucleus, or that the susceptibility of the endosymbiont to lose genes provided the selective pressure to create protein import machinery.

According to Gross et al., the first possibility represents the group of hypotheses called "insiders' models" [47, 48]. In this case, symbiont's genes serendipitously land near active promoters and/or reacquire existing promoters used by the host. Thus, genes would exist in duplicate until the system evolved relocation machinery to the proto-organelle. These insiders' models usually entail that to ensure the delivery and the assembly of host-encoded proteins in the newly established organelle, protein import machinery was needed. Some of the preexisting protein translocation apparatus of the endosymbiont appears to have been commandeered, including molecular chaperones, the signal peptidase, and

some components of the protein-targeting machinery [49-51]. On contrary, some descriptions emphasize the role of genetic innovations occurring on the host genome as well [52]. Alcock et al. [49] suggests that the components still encoded on the putatively minimally reduced α-proteobacterial endosymbiont genome were "tinkered" with by evolution to import nuclear-encoded proteins and thereby assumes that the α-proteobacterium had an important, if not preponderant, role it its conversion into an organelle.

In contrast, Gross et al. [47] offers a reversed view, the so-called "outsiders' model" of the evolution of the mitochondrial protein transport machinery. This model posits that genetic integration and the establishment of protein sorting system in mitochondria occurred in a step-wise evolutionary trajectory, with the host guiding molecular components first to the outer membrane (OM) of the endosymbiont, and then to the intermembrane space (IMS), inner membrane (IM), and finally to the organelle interior. Outsiders' model recognizes that by being held captive inside the host cell the mitochondrial ancestors were subject to the typical genomic "meltdown" universally observed among obligatory prokaryotic endosymbionts [53, 54]. Therefore the outsiders' model favors the idea that evolutionary novelties leading to the establishment of the organelle were mainly selected on the host chromosomes and contrasts with insiders' models that are united by the view that organelle protein sorting systems originated to target nuclear-encoded proteins into the endosymbiont interior [50, 52, 55, 56].

Only the future shows us which scenario is correct and it is also possible that the true lies somewhere between. Either way, from current phylogenetic data it is concluded that the protein translocases that drive protein import into mitochondria have no obvious counterparts in bacteria, making it likely that these machines were created de novo in the LUCA to all eukaryotes. The presence of similar translocase subunits in all eukaryotic genomes sequenced to date suggests

that all eukaryotes can be considered descendants of a single ancestor species that carried an ancestral proto-mitochondria.

2.2.1 The origin of targeting peptides

Most nuclear-encoded mitochondrial and hydrogenosomal precursors have an N-terminal presequence that is necessary at multiple translocation steps. How did these presequences get appended to hundreds of genes has been a difficult question to address. Detailed analyses using comparative genomics made clear the later stages of mitochondrial evolution, e.g. how escape of genes from the endosymbiont could enable transfer to the nucleus and integration in the genome, how adaptive rearrangements could allow gene expression and how exon shuffling and other recombination events could create a mitochondrial targeting sequence [57, 58]. However, these findings don't explain the very beginning of the evolution of the protein translocation machineries, in other words the co-evolution of mitochondrial membrane translocases and signal sequences. Lucattini et al. suggested that the major impetus for the development of the translocation machinery as a whole was likely to have relied on pre-adaptive features, literally basic amphipathic N-terminal segments, in many bacterial proteins that would serve as mitochondrial targeting information [59].

The above mentioned represents insiders' models that rely on an "adaptive" argument that if molecular machines were already present in the endosymbiont they could "somehow" be easily adapted to operate protein import inside the organelle. Thereby many insiders' models claim that evolution could be gradual, however they do not provide a clue for the order in which the molecular components forming the protein import machines were established. In other words, they do not incorporate a process-like description [47]. The ultimate root of the insiders' view is the signal hypothesis that initially intended to describe how a protein is targeted to the interior of endoplasmic reticulum [60]. Thereby it is assumed that the inherent function of a protein sorting system is to direct proteins to lumen of an organelle. These traditional models for the evolution of protein

topogenesis in mitochondria adhered to this idea and aimed at providing an insiders' description of organelle function and evolution.

Outsiders' hypothesis offers a different perspective to understanding the essential organization of a protein sorting system [47]. By applying a gradual outside-to-inside description of the evolution of protein sorting components that one can rationalize why precursors of proteins destined to the OM, IMS, and the mitochondrial carriers do not have N-terminal extensions (i.e. signal presequences). According to this idea it suggested that presequences appeared in a later stage to facilitate the insertion of proteins with single transmembrane domain into the IM by Tim23/Tim17 insertase, and only posteriori did they acquire the meaning of a signal. As a consequence it is suggested that presequences are actually signals for insertion into the IM and that the genuine signal for protein import into the matrix of mitochondria that emerged during evolution was the absence of an α-helical transmembrane domain following the presequence. Thus, outsiders' hypothesis claims that topogenic signals emerged from functional traits of the imported proteins.

2.2.2 Mitochondrial targeting peptides

Work done in yeast and other organisms shows that targeting sequences can be found at the N-terminus, C-terminus or internally in proteins destined for mitochondria. Many membrane proteins, for example, do not have the otherwise typical N-terminal targeting sequence that has been well studied and found on almost all soluble proteins and on many membrane proteins too. However, it remains a useful generalization to posit that mitochondrial proteins are designated by an N-terminal signal presequence. Although the nature of mitochondrial targeting sequences is more complex, the structural aspects in mitochondrial targeting sequences are conserved among all eukaryotes: for instance, mitochondrial proteins from animals or fungi are targeted to mitochondria of plants, and vice versa. In all cases, the targeting sequences studied feature:

- positively charged residues [61] and
- an ability to adopt amphipathic helix [62, 63].

In many cases, the targeting sequence also includes features that might assist its folding to a helical structure, and extensions that would allow for processing by the highly conserved inner membrane protease (IMP), mitochondrial intermediate peptidase (MIP) or MPP in the intermembrane space and matrix, respectively. Apart from the ability to adopt a positively charged amphipathic helix, the signal sequences recognized by MPP exhibit also following features:

- presence of an arginine residue in the position -2 and
- a hydrophobic residue in the position +1 relatively to the MPP cleavage site.

The same targeting sequence can serve to target proteins between all three classes of mitochondria-like organelles: hydrogenosomal sequences target proteins into mitosomes or mitochondria, and mitosomal sequences target proteins to mitochondria and hydrogenosomes [64, 65].

2.2.3 Transport of mitochondrial proteins

Protein uptake into a cell organelle is common and essential for all eukaryotic cells. This can occur basically by two mechanisms: co-translational import, which requires protein translocation to be tightly coupled to translation [66]; or post-translational import, in which protein synthesis is not mechanically linked to protein translocation. Generally, mitochondrial proteins are imported post-translationally. Since most of the functional studies to date have focused on the protein import machinery in the yeast *Saccharomyces cerevisiae*, we refer throughout the text to in yeast (unless stated otherwise).

The modular multi-subunit architecture of the protein import machinery provides the means to import and sort the many hundreds of proteins needed in mitochondria. Mitochondrial targeting sequences are recognized sequentially by a series of protein translocases [67, 68]. The protein translocases have a core translocation unit enhanced by one or more modules of distinct function, such as receptors, metaxins and accessory or peripheral proteins [50].

The main molecular machines for protein translocation across the mitochondrial outer and inner membrane are the translocase of outer and inner mitochondrial membrane (TOM and TIM), respectively (**Figure 6**). The TOM complex provides the gate for entry and is composed of several integral membrane protein components: Tom70 and Tom20 are receptor units [69] that recognize substrate proteins destined for import, with mitochondrial proteins bound by those receptor subunits subsequently released into a translocation channel composed of the core translocase components Tim40, Tom22, Tom7 and two small proteins, Tom6 and Tom5. Tom40 is probably a β-barrel protein, whereas Tom22, Tom5, Tom6 and Tom7 each have a single α-helical transmembrane segment that locks them tightly into position on Tom40 [70]. The hydrophobic surface of the amphipathic helix of the substrate presequence is recognized by Tom20 and a possible charged surface by Tom22 [62, 63, 71, 72]. As the driving force for precursor protein translocation is proposed the increasing affinity of the precursor proteins to theirs receptors in the order of their recognition ("affinity chain" hypothesis) [73].

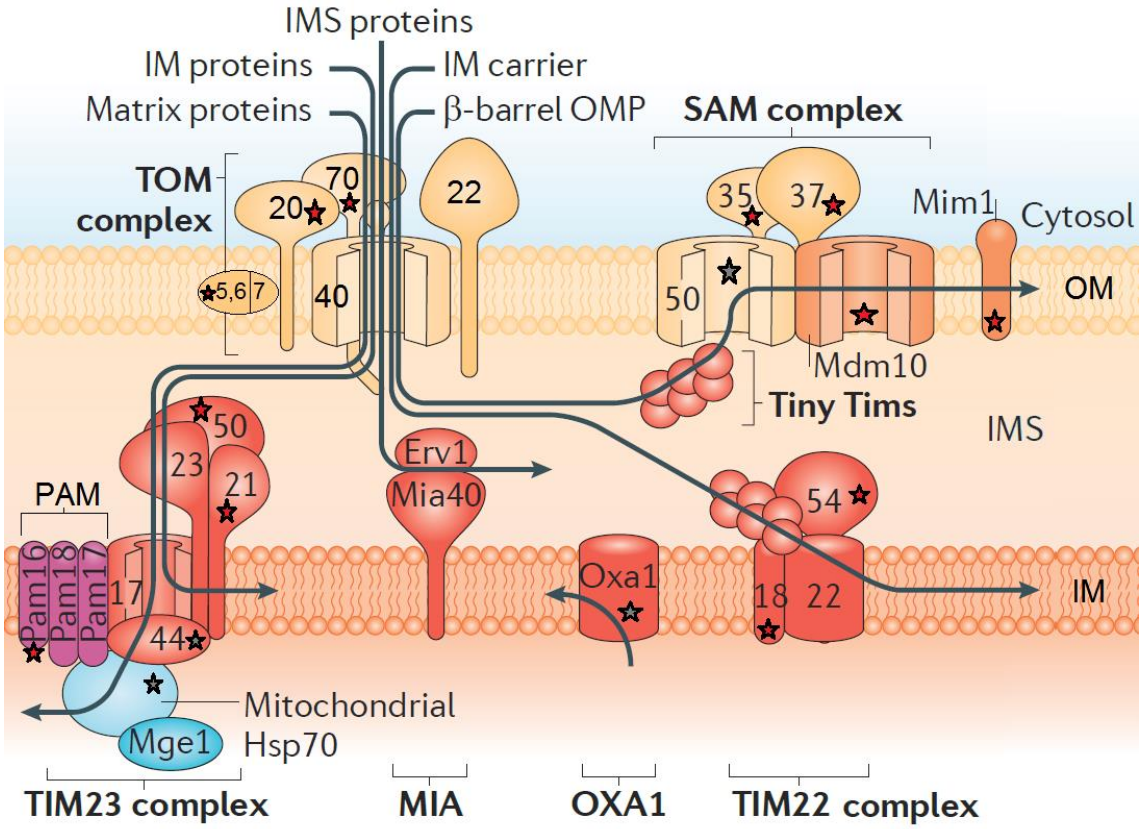


Figure 6. The protein import machinery in mitochondria. Precursor proteins are transferred across the outer membrane (OM) by the translocase of the outer membrane of mitochondria (TOM) complex (yellow) and are subsequently sorted to different locations within the organelle [74, 75]. The translocase of the inner membrane of mitochondria 23 (TIM23) complex (orange) acts in cooperation with the presequence-associated motor (PAM) complex (purple) and mediates protein transfer into the matrix [74]. The mitochondrial heat-shock protein 70 (Hsp70) provides the driving force within the PAM module and is regulated by its co-chaperones Pam18/16 and Mge1. In its PAM-free form, the TIM23 complex mediates the sorting of precursor proteins with a 'stop' signal into the inner membrane (IM). Oxa1 (which is related to the bacterial protein YidC [76]) facilitates protein insertion from the matrix into the inner membrane, whereas the TIM22 complex assembles carrier proteins with internal signal peptides into the inner membrane [74, 75]. The tiny Tim proteins guide β-barrel precursor proteins across the intermembrane space (IMS) to the sorting and assembly machinery (SAM) that catalyses integration into the outer membrane (OMP; outer membrane protein). The SAM components can also associate with Mdm10 and Mim1 to facilitate the biogenesis of outer membrane proteins. The mitochondrial IMS import and assembly machinery (MIA), which consists of Mia40 and Erv1, stimulates the import and assembly of IMS proteins with a characteristic cysteine motif [77]. Alternatively, proteins are first sorted into the inner membrane by the TIM23 translocase and subsequently released into the intermembrane space by proteolytic cleavage of the membrane anchor. Gray stars depict the functional homologs in bacteria and red stars mark the components that are present only in fungi and animals, which suggest that they might be modules added to the machinery relatively recently. Figure overtaken from [78] and adapted.

After passing through the channel of the TOM complex, substrate proteins can interact with one of the two distinct machines in the inner membrane (**Figure 6**). One of these, the TIM22 complex (translocase of the inner membrane of mitochondria) is built around the Tim22 subunit and binds only protein substrates destined for the inner membrane. The TIM22 machine is composed of four subunits embedded in the inner membrane and a peripheral set of ''tiny Tim'' subunits that shuttle to and from the TOM complex to collect substrates [67, 68, 79]. The translocation and insertion of inner membrane proteins by the TIM22 complex does not require adenosine triphosphate but depends on the electrochemical potential across the inner membrane [79]. Electrophysiological measurements demonstrate that the TIM22 complex contains pores that can flicker between different conformation states [80], which might reflect the movements the TIM22 subunit makes to assist substrate protein integration into the inner membrane.

The TIM23 complex is a distinct TIM complex built around a channel formed from Tim17, with this channel allowing for substrate entry to the mitochondrial matrix (**Figure 6**). Associated with the TIM23 complex, Tim50 is a receptor that guides protein substrates to bind the translocation channel [79] and thereby serves to regulate the opening and closing of the channel [81]. Tim21 interacts with the Tim17 subunit of the core translocase to assist in determining whether a bound substrate should be integrated into the IM or translocated through into the matrix [82, 83]. Translocation through the TIM23 complex is driven by a motor complex built around a mitochondrial Hsp70. The molecular chaperone Hsp70 is anchored to the membrane by proteins Pam16 and Pam18 [68, 79] and the peripheral inner membrane protein Tim44.

With regard to the outsiders' model for the evolution of protein translocation machines Tim23/Tim17 complex is of special interest, since this complex may act either as an insertase or as a translocase, depending on the absence or presence of the PAM module (presequence-associated motor) at the

mitochondrial matrix side, respectively. Thus, this situation illustrates that molecular machine might have evolved gradually from "outside to inside" [47]. An interesting fact is that Tim23 serves as an insertase for proteins containing a single transmembrane domain, which are abundant in the complex of the respiratory chain. In addition, in many taxa the subunits of MPP that cleaves off the presequences of import substrates are components of the respiratory chain complex III[1]. These correlations suggest that protein sorting in mitochondria evolved to support host-control of organelle oxidative phosphorylation. The fact that the host of the α-proteobacterial endosymbiont "reprogrammed" the prokaryotic "permeome" with host-derived solute transporters may be explained by two interpretations. One is the obvious host-control over the metabolic flow across the organelle membranes. The second might be a topological constrain to insert from outside molecular components derived from the endosymbiont, since the original prokaryotic transporters were assembled from inside.

2.3 M16 family peptidases

Metallopeptidase of family M16 are one of the most abundant classes of peptidases found in eukaryotes, bacteria and archaea [84]. These enzymes are zinc-ion-dependent peptidases with an HxxEH motif, which is an inverse of the classical metallopeptidase motif HExxH [85]. The physiological functions of some eukaryotic M16 members are as follows: degradation of physiologically important peptides such as insulin and amyloid β by insulysin [86, 87], removal of N-terminal targeting peptides from precursor proteins after their import into mitochondria and chloroplast by mitochondrial or stromal processing peptidase (MPP and SPP, respectively) [88, 89], degradation of cleaved-off targeting peptides by mitochondrial-targeting or stromal-targeting peptide peptidase (PreP) [90]. However, little is known about the intrinsic function of more than 2800

[1] Respiratory chain complex III is also known under the alternative term "bc_1 complex". Important for chapter 2.3.1.3 *Evolutionary context*.

bacterial sequences encoding prokaryotic M16 peptidases found in the peptidase database MEROPS [84].

M16 peptidases can degrade a wide variety of peptide substrates and lack selectivity for the amino acid sequences of substrates, although some preferences have been found. Insulysin cleaves a number of different peptides, such as insulin [91], transforming growth factor α [92], β-endorphin [93], amylin [94], and amyloid β [95]. MPP cleaves a variety of N-terminal signal sequences for mitochondrial transport [96, 97]; the resultant peptides are further degraded by PreP, which also lacks substrate specificity. Because of their broad substrate specificity and sequence selectivity, some M16 peptidases are rather designated as "peptidasomes" [98, 99]. Such a broad substrate specificity of M16 enzymes makes it difficult to elucidate their intrinsic substrates and biological roles.

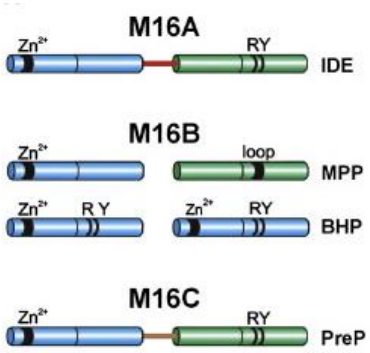


Figure 7. The domain organization of M16 family of peptidases. The key features are: the inverted Zn^{2+}-binding motif (HXXEH), the "peptidasome" motif RY and the "processing peptidase" GRL motif (loop). Figure taken from [100].

M16 peptidases are further categorized into three subfamilies (M16A, M16B, and M16C) according to their primary structures (**Figure 7**). Quaternary structures of M16 members show that enzymes in this family comprise four structurally similar domains. M16A and M16C peptidases (~100 kDa) contain four domains in a single chain [98, 101], whereas M16B enzymes (~50 kDa) form dimers in which each monomer contains two domains [88, 100]. For example, yeast MPP, a well-studied M16B enzyme, functions as a heterodimer consisting of

a β-subunit containing an HXXEH motif and an α-subunit without the motif [88]. The α subunit of the enzyme contains a glycine-rich loop (GRL) in the C-terminal domain that protrudes into the active site and is essential for catalysis [102].

Because of the importance of M16 proteases in the processing and degradation of many essential biological molecules, we need to increase our understanding of their function through structural studies of these enzymes from various sources, including bacteria (discussed in detail in chapter 2.3.3 *Bacterial M16B peptidases*) [100, 103]. The study of intrinsic function of M16 peptidases and structural factors determining their broad substrate specificity and sequence selectivity may also contribute to the discussion of the eukaryotic cell evolution and the origin of mitochondria, since these peptidases close the whole mitochondrial protein import machinery. Apart from that, these peptidases seem to be involved in serious human diseases, such as Friedreich 's ataxia [104]. There is also much interest in the possibility that insulysin may be a physiologically significant α-secretase, usefully degrading the amyloidogenic Alzheimer's β-peptide [105-107].

2.3.1 Mitochondrial processing peptidase

The activity of the mitochondrial processing peptidase was discovered in 1980 when Schatz et al. found that extracts of the mitochondrial matrix cleave off the presequences of in vitro synthesized mitochondrial precursor proteins [108]. Although it took about 20 years until three-dimensional crystal structure was resolved [88], many other structural and functional studies have been published, providing some clues to the understanding of MPP mechanism of function.

2.3.1.1 Structural features

MPP consists of two structurally similar but non-identical subunits, α and β, which are encoded by two separate genes. Each subunit is comprised of two domains of 210 amino acid residues each which share nearly identical topology. A conserved HxxEH zinc-binding motif, which appears only in the catalytic β-MPP

subunit, is located inside the enzyme cavity created by both subunits. Crystal structures of two different synthetic substrate peptides co-crystallized with mutant MPP showed the peptide bound in an extended conformation at the active site, forming a short series of β-sheet-like interactions with the β-sheets of β-MPP before proteolysis [88].

A common motif of almost all N-terminal signal presequences processed by MPP is an arginine residue located at position -2 with respect to the cleavage site (the R-2 motif) and additional arginine residues are usually located upstream in signal presequences. In the crystal structure, the R-2 residue was found to be interacting with E160 and D164, two conserved residues in the β-subunit which are close to the zinc-binding site (the R-2-binding motif). While the R-2 motif may be sufficient to indicate the cleavage site, it cannot be sufficient by itself for the initial substrate recognition, since MPP is able to specifically recognize a large variety of diverse mitochondrial signal presequences which generally have only low sequence similarity and vary in length.

2.3.1.2 Substrate recognition

Although it is known that both subunits are essential for MPP function [109-111], the mechanism of initial substrate recognition and the roles of each subunit in this process are still a matter of debate. Most studies have emphasized the importance of the α-subunit [112-114] in these processes and the ability of the signal presequences to form unstable α-helical amphipathic structures in hydrophobic environments, which seems to be important for the initial recognition of the presequence by MPP [115-117]. The most conserved part of all known α-MPPs is the glycine-rich loop (GRL; residues G^{285}GGSFSAGGPGKGMYS^{300} in the yeast α-MPP), which has been shown to be essential for substrate binding [102]. The GRL also seems to be the structural element where the initial interaction between MPP and the signal presequence occurs [118]. The most important role in this process may be a hydrophobic interaction between two residues on one side of the presequence α-helix (often in positions -4 and +1, with respect to the cleavage site)

with M298 and Y303 of the GRL. Since the GRL is situated at the entrance to the cleft formed by the MPP subunits, it is exposed to both the zinc-binding site and the substrate as it enters from outside the enzyme. Fluorescence resonance energy transfer experiments have shown that the C-terminus of the presequence interacts with the GRL following cleavage and points out of the enzyme cavity [119].

In summary, it seems likely that the ability of the presequence to adopt context-dependent conformations during different steps of MPP's action is a basic requirement for substrate recognition and processing. Although the MPP crystal structure was published more than a decade ago [88], very few studies have addressed the detailed operation of MPP. To date, the substrate binding and cleavage mechanism in the MPP active site has been described in detail [120] and a mechanism of substrate recognition by the GRL has been outlined [118] however the mechanism of substrate translocation from the GRL to the MPP active site has not yet been investigated.

2.3.1.3 Evolutionary context

In autotrophic and heterotrophic organisms MPP is located in different submitochondrial compartments: the enzyme from yeast and mammals forms a heterodimer in the mitochondrial matrix while in plants the two MPP subunits substitute the "Core I" and "Core II" proteins of the bc_1 complex of the inner mitochondrial membrane (and thus fulfill two physiological functions simultaneously) [121]. However, the core subunits of the bc_1 complex of mammals (**Figure 8**) exhibit sequence similarity to the subunits of MPP but lack a complete zinc-binding motif[2] and are proteolytically inactive. Interestingly, some processing activity of the bovine bc_1 complex can be activated in vitro by detergent treatment or by the combination of purified Core I and Core II proteins after overexpression in *E. coli* [122, 123]. In vivo the bovine bc_1 complex is most likely proteolytically

[2] However, a sequence of $Y_{57}XXE_{60}H_{61}(X)_{76}E_{137}$, similar to the zinc-binding motif HXXEH in the β-MPP subunit, is present in the Core I protein of the bovine complex.

inactive because a peptide generated by proteolytic cleavage of the Rieske iron–sulfur subunit is attached to the active site of Core II subunit, as revealed by crystallization of the bc_1 complex [124].

Mammalian and plant organization of MPP/bc_1 complexes represents "boundary states". However, examples of intermediate situations exist among fungus species:

- The β-MPP subunit is identical to the Core I and is present in both the matrix and inner membrane, but α-MPP is present in the matrix only. Such organization was found in *Neurospora crassa* [125, 126] and *Dictyostelium discoideum* [127].
- In *Blastocladiella emersonii*, β-MPP is found in both the matrix and the inner membrane, but α-MPP is found in the inner membrane [128].

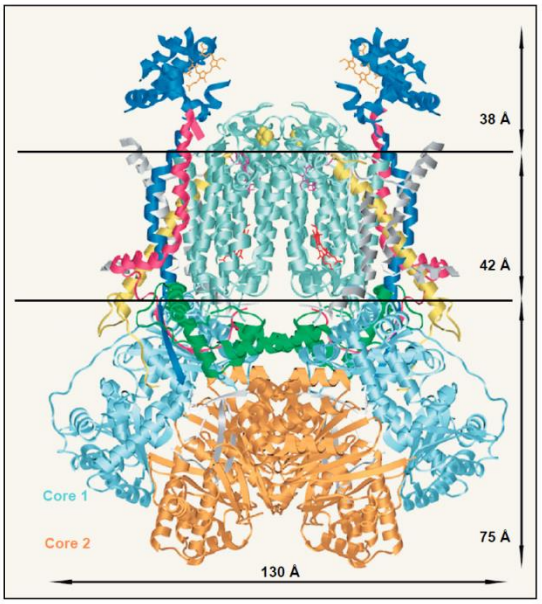


Figure 8. Three-dimensional crystal structure of the bovine bc_1 complex. Horizontal lines mark the outer and inner side of the inner mitochondrial membrane. The membrane part of bc_1 complex is 42 Å thick with 12 transmembrane helixes, and the intermembrane and matrix parts extend 38 Å and 75 Å into the corresponding space. Core I and Core II subunits are highlighted in cyan and orange color, respectively. Figure taken from [124] and adapted.

It has been suggested that the progenitor of MPP was probably a monomeric α-proteobacterial peptidase, similar to the contemporary *Rickettsia prowazekii* processing peptidase (RPP) [129]. During the evolution of mitochondria, gene duplication and subunit specialization gave rise to the heterodimeric MPP. Bifunctional role of plant bc_1 complex then may represent a situation that was once common to all mitochondrion progenitors [122].

According to the outsiders' model for the evolution of mitochondrial protein import machinery, the MPP-bc_1 correlations suggest that protein sorting in mitochondria evolved to support the host-control of organelle oxidative phosphorylation [47].

2.3.2 Hydrogenosomal processing peptidase

Hydrogenosomes are highly reduced versions of mitochondria that are found in diverse parasitic unicellular eukaryotes inhabiting intracellular niches [37]. The hydrogenosomes found in human parasite *Trichomonas vaginalis* lack a genome[3] so all of their proteins are encoded by the nuclear genome and must be imported [37].

Some hydrogenosomal proteins have N-terminal extensions that are reminiscent of the presequences that direct proteins into mitochondria and they contain distinguishable cleavage motifs [25]. This suggested that the *Trichomonas* organelles may also contain an MPP-like enzyme. A single gene for a β-MPP homologue (20.9% identity and 42.9% similarity to *S. cerevisiae* β-MPP) was identified in the genome of *T. vaginalis* [130]. In this case, functional data were presented suggesting that the hydrogenosomal processing peptidase (β-HPP) functioned as a homodimeric enzyme [130]. No α-MPP homologue was detected, although a glycine-rich-loop-like protein (GRLP) , that shares a limited similarity

[3] However, hydrogenosomes possessing DNA also exist [20].

with the GRL of α-MPP, was located to *T. vaginalis* hydrogenosomes. However, GRLP was reported not to stimulate β-HPP activity in vitro [130].

However, recently we have shown that the HPP is fully active only upon heterodimerization of an α and β subunit, like classical MPP [131]. The processing activity was demonstrated as a shift in the substrate gel mobility and thus the structural information was just indirect. Although such approach is sufficient for the purpose of this publication, deeper insight into HPP structure was missing. Thus far, the three-dimensional structure or any other description of the structure of HPP has not been presented.

2.3.2.1 Evolutionary context

Mitosomes are another example of highly reduced version of mitochondria and similarly to hydrogenosomes, a gene coding for a putative processing peptidase has been found in the genome of *Giardia intestinalis* (another example of serious eukaryotic human intracellular parasite, member of the diplomonads) [132]. The gene product "*Giardia* processing peptidase"(GPP) has been shown to localize in mitosomes [133]. The primary structure of GPP is highly divergent from mitochondrial homologues, with only 13.1% identity and 29.7% similarity to the yeast β-MPP. It has been shown that GPP functions as a monomer consisting of a single β-MPP homologue [131].

Since it is widely accepted that the progenitor of mitochondria was of α-proteobacterial origin, it is also assumed that the progenitor of MPP was probably a monomeric α-proteobacterial peptidase [129]. Thus, the single subunit architecture of GPP and also HPP (as originally suggested by Brown et al. [130]) could reflect retention of the ancestral form of organization. However, in the light of our result it seems to be more likely that both peptidases are products of the reductive evolution from the classical MPP heterodimer [131]. Moreover, phylogenetic analysis supports the hypothesis that GPP, β-HPP and β-MPP share a common origin and suggests that all MPP-like proteins arose by a primordial

gene duplication of single subunit α-proteobacterial peptidase. While HPP represents an intermediate stage between GPP and MPP, GPP provides a striking example of strong reductive evolution from a heterodimeric to a monomeric enzyme, with properties resembling the putative ancestral α-proteobacterial enzyme.

The identification of other mitochondrial remnant organelles in amitochondrial protists such as *Entamoeba histolytica* [134, 135], *Trachipleistophora hominis* [136] and *Trypanosoma brucei* [137] suggest that the eukaryotic amitochondrial state is not a primitive condition but is rather the result of reductive evolution.

2.3.3 Bacterial M16B peptidases

Bioinformatic analysis has pointed to the existence of prokaryotic members of the M16B subfamily, although their biological functions are unknown. The MEROPS database [84] subdivides them further into four groups (**Table 1**), each comprising a single ~500 residue domain, although a few members of M16B.UPB have the two domains fused, analogically to M16A/C class. Some of them contain typical zinc-binding motif and all have reasonable homology (up to 30% identity) with their eukaryotic M16 counterparts. However it should be noted that they do not contain the long GRL, a typical trait of eukaryotic M16B peptidases, but some groups contain an R/Y pair (i.e. RY motif) conserved in the C-terminal domain. The RY motif has been shown to interact with substrate residues in the vicinity of its cleavage site whereby posits the substrate to the proper position required for the subsequent proteolysis.

Group	Zn motif	RY motif	Examples
M16B.016	Yes	Yes	Homodimers (peptidasomes): • BHP [100, 138]
M16B.014	Yes	No	Heterodimers: • BH2392 [100]
M16B.UNB	No	Yes	Heterodimers: • SPH2682 [139] • TTHA1265 [100] • BH2393 [100]
M16B.UPB	Yes	No	Heterodimers: • SPH2681 [139] • TTHA1264 [103]

Table 1. Organization of the prokaryotic M16B family. M16B.016 members are homodimers (green). M16B.UNB members constitute heterodimers either with M16B.014 (blue) or M16B.UPB (orange) family members. All prokaryotic members of M16B family don't possess GRL.

Until recently, prokaryotic M16B peptidases have been suggested to be active in monomeric form based on the enzymology of *Rickettsia prowazekii* peptidase RPP [129] and *Bacillus halodurans* peptidase ppBH4 [138]. However, in the last couple of years a three-dimensional structure has been resolved also for some prokaryotic members of the M16B family. Below continues the description of all so far biochemically and/or structurally characterized prokaryotic M16 peptidases.

2.3.3.1 Rickettsia prowazekii peptidase

The obligate intracellular parasitic α-proteobacteria rickettsiae are more closely related to mitochondria than any other microbes analyzed on the genome level to date [9]. Interestingly, *Rickettsia prowazekii* gene 219 encodes a putative peptidase ("*Rickettsia* Processing Peptidase" - RPP) that is similar to MPPs (42% with yeast β-MPP). The structural characteristics (but not three-dimensional structure) and biochemical activities of RPP were analyzed by Kitada et al [129]. The RPP primary structure resembles those of both of the MPP subunits, since the N- and C-terminal regions of RPP were similar to the N-terminus domain of the β-MPP

(including the zinc-binding motif) and C-terminus domain of the α-MPP subunit, respectively. The C-terminal domain or RPP doesn't possess GRL or RY motif.

Since RPP contains the catalytic zinc-binding motif, the proteolytic activity of RPP was examined. Peptides containing basic amino acid residues produced detectable amount of fragments following incubation with RPP, whereas cleavage of neutral and acidic peptides by RPP was undetectable. Mass spectrometry was used to show that RPP preferably hydrolyzed peptide bond before hydrophobic residue and sometimes attacked sites beside basic residues, which MPP also prefer. RPP was further investigated whether could cleave mitochondria-targeting presequences. RPP was shown to process a 16-residue long peptide derived from yeast MDH signal presequence (residues 2-17) at the same cleavage site as MPP does. Among the eight peptides tested, RPP attacked five of them at one (corresponding to MPP processing site) or more sites.

The processing activity of RPP was further tested with two non-cleavable peptides in the excess of α-MPP and β-MPP subunits. Namely, with 27-residues and 40-residues long peptide derived from mouse MDH and bovine adrenodoxin (ADX) signal presequence (residues 2-28 and 18-57), respectively. Interestingly, a stoichiometric mixture of RPP and yeast β-MPP processed mouse MDH, whereas a similar mixture of RPP and yeast α-MPP did not. On the other hand, bovine ADX was cleaved in neither case. Yeast β-MPP alone did not show any processing activity and thus functional association between RPP and β-MPP was established. However, the stable RPP-β-MPP interaction was not demonstrated by β-MPP pull-down assay using the His_6 tag of RPP and affinity beads.

Authors concluded that RPP behaved as a regulatory subunit toward mouse MDH processing, similarly to α-MPP subunit. To compare regulatory function of both proteins, a mutant form of α-MPP lacking GRL was prepared (Δ287-294). Authors showed that ΔGRL MPP was able to cleave mouse MDH, albeit with a lower processing activity than that of wild-type MPP, but showed inefficient in bovine ADX processing. Author state that a pull-down assay of

ΔGRL α-MPP by β-MPP previously revealed a stoichiometric association of both subunit [102][4]. Thus, Kitada et al. concluded that the GRL region is required only for the cleavage of long signal presequences but does not influence the association between MPP subunits.

However, in this part, the work of Kitada et al. should be considered with caution, since we have found numerous mistakes throughout the whole publicatio. In addition, the main conclusion is based on experiments with only one signal presequence (mouse MDH).

The physiological function of RPP in vivo remains unknown. However, its function may be derived from the *ymxG* gene product, a M16 family protease from eubacterium *Bacillus subtilis* with high homology with MPPs (51% of identity with β-MPP subunit) [140] and from MPP-like proteins from some parasitic eubacteria. Analyses of Δ*ymxG* mutant strain revealed specific stimulation of the production of subtilisin (AprG), the major serine protease secreted by *B. subtilis* [141]. This phenomenon appears to arise through negative regulation of *aprE* gene expression by YmxG rather than through the lack of YmxG proteolytic activity. Considering the homology of RPP to YmxG and the RPP peptidase and regulatory activities, RPP may play a key role in regulating protein expression through its protease activity. Even though the RPP functions in vivo remain unknown, the fact that the eubacterial MPP-like protein YmxG is not essential for viability or cell growth [141] is interesting when the origin of these preprotein processing enzymes is considered according to endosymbiotic evolution of mitochondria.

According to Kitada et al. [129], BLAST searches of prokaryotic genomes revealed nearly 500 homologues of yeast β-MPP subunit, whereas proteins bearing

[4] In this case, authors state that they purified yeast ΔGRL MPP mutant using a His_6 tag of β-MPP. However, they reffer to previous publication [102] where ΔGRL MPP was purified using **rat** β-MPP and yeast His_6 tagged α-MPP.

GRLs were not found in bacterial genomes[5]. Considering the genomic distributions of MPP-like genes, and since it is widely accepted that mitochondria originated from the α-proteobacterial order *Rickettsiales*, MPP is unlikely to have originated from proteins in the ancestor host cells and is most likely to have arisen from RPP-like progenitor in a parasitic bacterium. During the endosymbiotic evolution of mitochondria, the gene encoding the progenitor of MPP, which might be nonessential, similarly *ymxG* in *B. subtilis*, could be transferred from the endosymbiont to the host cell. At his stage, the nonessential gene conversion presumably might succeed during mitochondrial evolution due to the maintained viability of the imperfect symbiotic organelle. Subsequently, during the evolution from endosymbiont to mitochondria, the progenitor gene was duplicated and the proteins were converted into two distinct components of the processing enzyme, which are now α- and β-MPP subunits.

2.3.3.2 Bacillus halodurans peptidase

The crystal structure of the BH2405 gene product of *Bacillus halodurans* strain C-125 strain ("*Bacillus halodurans* Peptidase" - BHP) has been resolved recently and it has been shown, in contrast to previous report[6] [138], that BHP forms homodimer [100]. BHP is classified as a member of M16B.016 subfamily and each of its subunits bears zinc-binding and RY motif. Although BHP resembles its eukaryotic M16B orthologs, the substrate binding induces complete domain closure (**Figure 9**) to a state that closely resembles the M16A/C class, consistent with its biochemical activity in which it functions as a peptidasome rather than a processing peptidase. In addition, the domain closure is required for catalytic activity of BHP.

[5] In the context of further discussion it is worth noting that Kitada et al. meant here the typical "full-length" GRL of yeast α-MPP. However, in three-dimensional structures of some bacterial MPP-like peptidases it is possible to trace a small loop that we call the "embryonal" GRL.
[6] In this publication the peptidase is designated as ppBH4. It is biochemical characterization study.

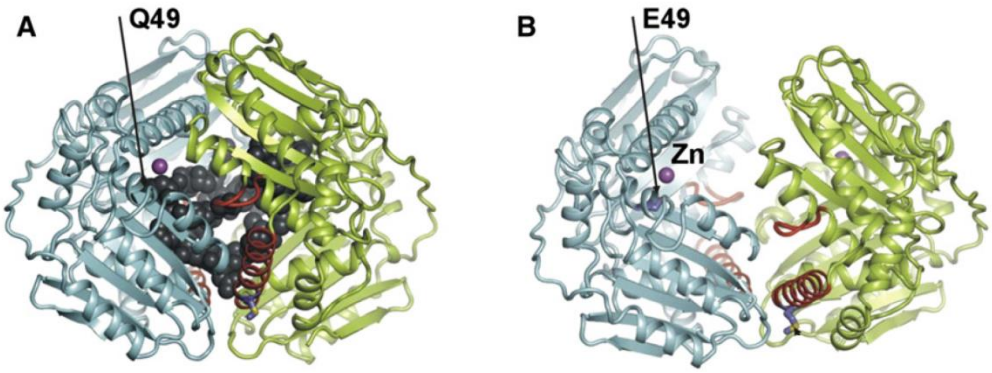


Figure 9. Closed and open form of BHP. Panel A shows the E49Q mutant with insulin-B bound in its active site, while panel B shows the wild-type form of BHP without substrate. Figure taken from [100] and adapted.

In addition to BHP, the genome of *B. halodurans* contains two adjacent genes coding for M16B.UNB and M16B.014 subfamily members. Aleshin et al. [100] proposed that members of subfamilies M16B.014 and M16B.UPB (containing zinc-binding but not RY motif) form heterodimers with members of the M16B.UNB members, which lack zinc-binding motif but contain RY motif (**Table 1**). Both subunits are typically encoded by adjacent genes. A second example of such an organization might be then M16 peptidase from *Thermus thermophilus* or *Sphingomonas sp*.

2.3.3.3 Thermus thermophilus peptidase

The crystal structure of a putative peptidase from *T. thermophiles* strain HB8 (TTHA1264) was solved as the first structure of bacterial putative peptidase belonging to M16 family [103]. The crystal structure contains two identical subunits homologous to β-MPP subunit. Surprisingly, while the zinc-binding motif (**Figure 10**) is present in both subunit contains, RY motif is missing.

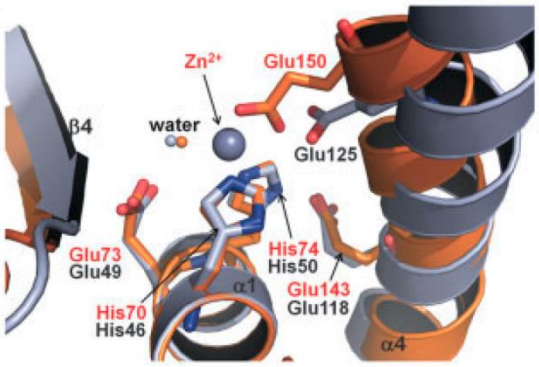


Figure 10. The zinc-binding motif and the surrounding residues of TTHA1264 (gray) and yeast β-MPP (orange). The residues in zinc-binding motifs are represented as stick models. The zinc ion (gray) and the catalytic water molecule (orange) of β-MPP are shown as spheres. A water molecule of TTHA1264 corresponding to the catalytic water is represented as a gray sphere. Figure taken from [103] and adapted.

This protein is monomeric in solution, and although it crystallized as a homodimer, the authors proposed that the dimer was an artifact of crystallization. Therefore, the authors conclude that TTHA1264 may function as a monomer like other bacterial homologues, such as ppBH4 [138] and RPP [129].

On contrary, Aleshin et al. [100] later predicted that this peptidase lacks intrinsic proteolytic activity and as a member of M16B.UPB subgroup must form a heterodimer with a subunit containing RY motif to be functional. Indeed, a gene for such subunit exists in *T. thermophiles* genome – according to the sequence, the predicted product of this gene, a protein designated as TTHA1265, contains the RY motif and also the rest of GRL. Furthermore, TTHA1265 belongs to M16B.UNB subgroup (**Table 1**).

Thus, although this structure is not probably representing the native state of TTHA and is very likely artificial, the "embryonal" form of the GRL is still visible in the structure. In this case, further investigation of the subunit composition is needed.

2.3.3.4 Sphingomonas sp. peptidase

The three-dimensional structure of a putative peptidase from α-proteobacterium *Sphingomonas sp.*[7] strain A1 was solved recently using X-ray crystallography (SPH) [139]. The peptidase was found to be heterodimeric in nature and composes of two subunits (SPH2681 and 2682). The two genes coding for each subunit assemble into a single operon in the bacterial genome. SPH 2681 subunit contains the HXXEH zinc-binding motif and SPH2682 conserves an RY motif commonly found in the C-terminal half of M16 enzymes. Thus, the first is homologous with mitochondrial β-MPP subunit and the later to α-MPP (although doesn't contain GRL).

SPH2681 was found to associate with SPH2682, forming a heterosubunit enzyme with peptidase activity, while SPH2681 alone exhibited no enzymatic activity. X-ray crystallography of the SPH complex revealed two conformations (open and close) within the same crystal. Compared with the closed form, the open form contains the two subunits rotated away from each other by ~8° (from 10° for the open to 2° for the closed form; **Figure 11**), increasing the distance between the zinc ion and RY motif by up to 8 Å. In addition, many hydrogen bonds and salt bridges are formed or broken on change between the conformations of the heterodimers, suggesting that subunit dynamics is a prerequisite for catalysis. Although open and closed conformations are not unknown among eukaryotic members of the M16 family, this is the first report on both conformation of the same M16 peptidase. Moreover, SPH is a prokaryotic M16 member and thus this study provides unique insights into the general

[7] The genus *Sphingomonas* was created by Yabuuchi et al. [142] to accommodate strictly aerobic and chemoheterotrophic gram-negative bacteria that contain glycosphingolipids (GSLs) as cell envelope components. Since then, large numbers of phenotypically and phylogenetically similar strains from various environments have been added to this genus and described as novel species of this genus. As a result, the genus *Sphingomonas* has come to encompass a relatively broad range of species with respect to physiology, genetics and ecology (including a wide variety of pathogens and free-living bacteria).

proteolytic mechanism of M16 proteases and also provides clues to the possible MPP evolutionary scenario, as will be discussed later.

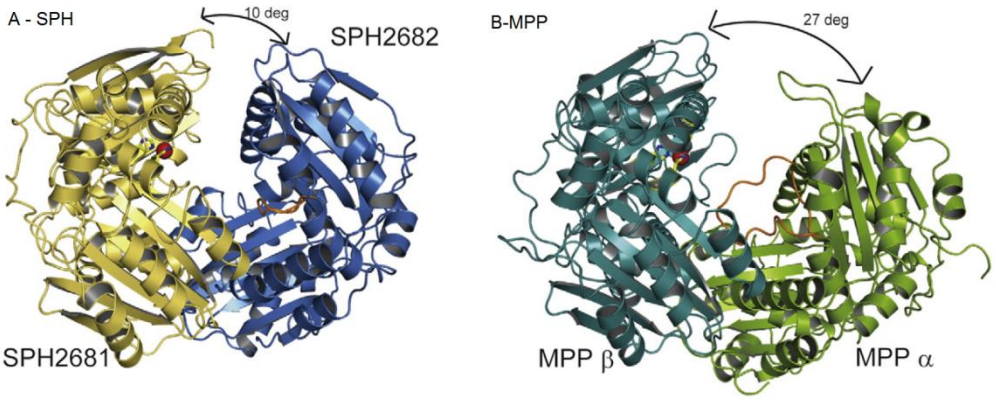


Figure 11. Comparison of subunit architecture between SPH and yeast MPP. Panel A shows the open dimer of SPH2681/SPH2682 and Panel B shows the dimer of MPP. The red sphere represents zinc ion. When the subunits are defined as planes, the angle between SPH2681 and SPH2682 is about 10° for the open form, whereas the corresponding angle between MPP α- and β-subunit is about 27°. Figure taken from [139].

3. A I M S

3.1 The role of GRL in MPP structure and function

The whole process of substrate translocation from the GRL to the MPP active site is complex and involves interactions of larger regions rather than single amino acids and the experimental approach (using, for instance, amino acid point mutations) is problematic. Thus, the aim in this part is to determine the precise role of the GRL in different steps of the substrate processing using molecular dynamics (MD) simulations, which may provide new insights into MPP substrate recognition and processing mechanisms which are difficult to acquire experimentally. Specifically, the aims in this part are as follows:

1. Construct and validate a model of MPP in interaction with its substrate in the place of GRL and in its active site.
2. Employ targeted MD simulation to study the process of substrate translocation from GRL to MPP active site.
3. Using non-restrained MD simulation study in detail the selected moments (i.e. MPP-substrate interaction and mutual conformation) along the substrate translocation trajectory.
4. Construct a GRL-deleted model of MPP and using non-restrained MD simulation study the effect of this mutation on the overall structure of MPP, both in the presence and absence of the substrate in its active site.

3.2 Structural features of HPP

Since the information about the structure of HPP are very limited, in other words we have just an indirect evidence that HPP forms a dimer [131], the aims in this part are as follows:

1. Construct and validate the homology model of HPP.

2. Optimize expression and purification conditions for HPP. Employ pull-down assay to verify subunit interaction.

3. Perform crystallization experiments and eventually solve the three-dimensional structure of HPP.

4. Study the nature and the dynamic of the conformation changes of HPP in the presence or absence of its substrate using following approaches:

    a. Biological small-angle X-ray scattering (bio-SAXS)

    b. Tryptophan fluorescence

    c. Hydrogen/deuterium exchange coupled with mass spectrometry

    d. Cross-linking techniques coupled with mass spectrometry

3.3 Evolutionary path of MPP and its GRL

GRL of MPP is the place of the substrate recognition and very likely participates also in substrate translocation to the MPP active site. Although MPP evolved very likely from an α-proteobacterial peptidase, the contemporary bacterial peptidases lack GRL. However, they do contain an "embryonal" form of GRL. On contrary, HPP is a hydrogenosomal evolutionary successor of MPP that probably still contains GRL, however only in the modified form.

Hence, the aim in this part is as follows:

1. Work out the evolutionary scenario for GRL, in the context of the MPP evolution from a bacterial peptidase to its reduced forms found in hydrogenosomes and mitosomes.

4. MATERIAL AND METHODS

4.1 Material

4.1.1 Cloning vectors and constructs

pJAKO – A cloning vector designed in our laboratory (J. Janata, J. Kopecký). It is based on pBluescript II KS+ (Stratagene) and bears modified poly-linker with *NdeI* and *NcoI* restriction sites. It bears a gene for ampicillin resistance.

pET vectors (Novagen) – A set of vectors for heterologous protein expression in *E. coli* cells. pET42b was the most frequently used one; it bears a gene for kanamycin resistance.

pET and pJAKO derived constructs – A number of constructs coding for various versions of α- and β-HPP subunits prepared during the process of conditions optimization for expression and for further experiments. See **Table 2**.

Construct Name	Vector Name	Clonning Sites	Res.	Notes
α-HPP subunit vectors				
JATV	pJAKO	NcoI, XhoI	amp	Stop codon Yes; 2 artificial residues at N-term; extra NdeI in gene
PATV	pET28b	NcoI, XhoI	kan	Stop codon Yes; 2 artificial residues at N-term; extra NdeI in gene
JATVD	pJAKO	NdeI, XhoI	amp	Stop codon Yes; 2 artificial residues at N-term removed; extra NdeI in gene
JATVE	pJAKO	NdeI, XhoI	amp	Stop codon Yes; 2 artificial residues at N-term removed; extra NdeI in gene removed
PATVE	pET28Amp	NdeI, XhoI	amp	Stop codo Yes; His-tag at N-term; 2 artificial residues at N-term removed; extra NdeI in gene removed
PATVN	pET42b	NcoI, XhoI	kan	Stop codon Yes; His-tag at N-term; 2 artificial residues at N-term; extra NdeI in gene
JATVB	pJAKO	NdeI, BamHI	amp	WT; Stop codon > BamHI
PATVB	**pET42b**	**NdeI, XhoI**	**kan**	**His-tag at C-term; extra NdeI in gene**
PATVB-W236Y	-	-	-	-
PATVB-W236F	-	-	-	-
PATVB-W236Y-F256W	-	-	-	-
PATVB-W236Y-F261W	-	-	-	-
PATVB-W236Y-F267W	-	-	-	-
PATVB-W236Y-F271W	-	-	-	-
β-HPP subunit vectors				
JBTV	pJAKO	EcoRI, XhoI	amp	WT; Stop codon Yes
PBTV	**pET42b**	**NdeI, XhoI**	**kan**	**WT; Stop codon Yes**

PBTV-E56Q	pET42b	NdeI, XhoI	kan	WT with E56Q; Stop codon Yes
JBTVB	pJAKO	NdeI, BamHI	amp	WT; Stop codon > BamHI
PBTVB	pET42b	NdeI, XhoI	kan	His-tag at C-term
PBTVN	pET28b	NdeI, XhoI	kan	Stop codon Yes; His-tag at N-term
Adenylate-kinase vectors				
GAKTV	pGEM	NdeI, XhoI	amp	
PAKTVSC	**pET42b**	**NdeI, XhoI**	**kan**	**His-tag at C-term**

Table 2. The list of all prepared vectors used in experiments. The first letters "P" and "J" in vector names designate pET and pJAKO parent vectors, respectively. Second letters "A" and "B" designate α and β-subunit, respectively. "TV" refers to *Trichomonas vaginalis*. Ending letter "B" designates His-tag at C-term (BamHI site used for cloning), if present. In bold are marked vectors proved to produce pure and stable HPP dimer and used in majority of experiments (crystallization, SAXS, H/D exchange, cross-linking).

pGroESEL (DuPont) - Vector bearing sequences coding for GroES and GroEL chaperonins used for heterologous protein expression in *E. coli*. It bears a gene for chloramphenicol resistance.

pET Duet (Novagen) – pET Duet-1 was used for tandem expression of α- and β-HPP subunits (hence HPP-Duet). Here, α-HPP subunit is in its native form while β-HPP subunit has attached His-tag at N-terminus.

4.1.2 Bacterial strains

JM109 (Stratagene) – E. coli strain used as a recipient for plasmid DNA transformation for the purposes of DNA overproduction.

- e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17($rK^- mK^+$) supE44 relA1 Δ(lac-proAB) [F′ traD36 proAB lacIqZΔM15]

BL21 (DE3) (Novagen) – *E. coli* strain used for heterologous protein expression. It bears gene for T7 RNA polymerase (DE3) under control of *lac* operator.

- F- ompT [lon] hsdSB ($rB^- mB^-$) gal dcm

Rosetta (Novagen) - Rosetta host strains are BL21 *lacZY* derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*.

4.1.3 Material

Ni-charged beads **-** Ni-charged beads (Ni-Sepharose High Performance GE Healthcare) were used for metal affinity purification of histidine-tagged proteins. The matrix consists of 34 μm beads of highly cross-linked agarose, to which a chelating group has been couples. This chelating group has then been charged with Ni^{2+} ions.

Protein standard **–** PageRuler Prestained Protein Ladder Plus (Fermentas) was used for SDS-PAGE protein electrophoresis.

DNA standard – 1 kb Plus DNA Ladder (Gibco) was used for DNA agarose electrophoresis.

4.2 DNA manipulations

4.2.1 General DNA techniques

DNA transformation – E. coli strain JM109 was used for overproduction of plasmid DNA. Heatshock method was used for cell transformation, as described previously [143].

DNA isolation - Following commercially available kits were used for DNA isolation from bacterial cell cultures, purification or buffer exchange:

- DNA Purification System Midipreps (Promega)
- DNA Purification System SV Minipreps (Promega)
- SV Gel and PCR Clean-Up System (Promega)
- DNA Clean-Up System (Promega)

Routine agarose electrophoresis was used for DNA quality controls and concentration estimations, as described previously [143]. In addition, NanoDrop (Thermo Scientific) measurements were employed for the same purposes.

DNA restriction & ligation – For DNA restriction were used commercially available restriction endonucleases and ligases (New England Biolabs) according to manufacturer general recommendations.

4.2.2 Site-directed mutagenesis

QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) was used for introduction of point mutations in the studied proteins, according to manufacturer recommendations.

Specifically, the method was used for removing native tryptophan residues and introduction of reporter ones in the sequence of GRL. Method was also used for preparation of inactive form of β-HPP subunit with E56Q mutation in its active site.

4.2.3 pET Duet expression system

pET Duet system (pET Duet-1, Novagen) was used for tandem expression of α- and β-HPP subunits in BL21(DE3) *E. coli* strain. Vector was prepared according to the manufacturer recommendation.

4.3 Protein expression and purification

All protocols for protein expression, purification, activity, concentration, storage and stability described in this chapter were optimized especially for HPP and adenylate-kinase (AK) as its substrate. The protocols for purification of MPP were elaborated previously [118].

4.3.1 Protein expression

For protein expression was used pET expression system (Novagen). In this system, the target gene is under control of T7 promotor. After addition of the IPTG as a inductor, the transcription of T7-RNA polymerase is started and T7 polymerase, in turn, starts to transcribe the target gene.

The wild-type and mutant forms of α- and β-HPP subunits were produced separately in *E. coli* BL21 (DE3) strain transformed with appropriate constructs and co-transformed with pGroESL (vector bearing sequences coding for GroES and GroEL chaperonins). Transformed *E. coli* cells were grown overnight on LB plates containing 30 μg.ml of kanamycin and 34 μg.ml of chloramphenicol. Subsequently, the cell were washed and transferred into liquid LB medium (with 30 μg.ml of kanamycin and 34 μg.ml of chloramphenicol). The cultivation was performed at 37 °C and 200 rpm for several hours. When the culture has reached OD_{600} = 0.6-0.7, it was induced with IPTG (final concentration 100 μM) and cultivated overnight (13 - 15 h) at 24 °C and 200 rpm.

Adenylate-kinase (AK) was produced in soluble form in *E. coli* BL21 (DE3) strain transformed with PAKTVSC construct. Similarly to production of HPP subunits, the transformed cells were grown on LB plates and subsequently transferred into liquid LB medium (containing 30 μg.ml of kanamycin). The culture was incubated at 37 °C and 200 rpm until it reached OD_{600} = 0.6-.07, when protein expression was inducted by addition of IPTG (final concentration 400 μM). The cultivation continued under the same conditions for one hour and then the cells were collected.

4.3.2 Metal affinity purification

The wild-type and mutant forms of HPP dimer was isolated using metal affinity chromatography. For one purification run, 1.5 l and 0.25 l of cell culture with expressed α- and β-HPP subunits was used, respectively. The cells were collected by centrifugation at 4 °C and 4000 g for 10 min and the cell pellet was, in turn, resuspended in 40 ml of resuspension buffer containing 25 mM HEPES, 20 mM NaCl, 5% glycerol, pH 7.5. Cell suspension was sonicated for 60 s on ice to disrupt the cell walls and to release expressed proteins and followed by centrifugation at 4 °C and 16000 g for 20 min. The pellet containing cell walls, chromosomal DNA and insoluble proteins was discarded and the supernatant with soluble proteins was used for further processing.

In the next step of purification, 1 ml of fresh (or 2 ml of used) Ni-charged beads was added to the supernatant, the suspension was gently mixed and placed on ice for 5 min. The suspension was shortly centrifuged (3 000 g, 3 s) to separate the beads with bound proteins from the supernatant. The beads were transferred into a column and washed in the steps with 15 and 5 ml of pre-cooled resuspension buffer. Further, the beads were washed in two steps with 10 and 5 ml of pre-cooled washing buffer containing 25 mM HEPES, 20 mM NaCl, 20% glycerol, 100 mM imidazole, pH 7.5. The 1-ml fractions were collected. The bound proteins were released from the beads by two-step washing with washing buffer with 250 mM imidazole. Five 1-ml and seven 5-ml fractions were collected. All collected fractions were visually checked on SDS-PAGE (**Figure 13**), fractions contacting proper HPP dimer were combined and subjected to following concentration. To obtain better yield of HPP, the supernatant was again mixed with new Ni-charged beads and the purification procedure was repeated.

The metal affinity purification of AK was performed using Hi-Trap Chelating columns (GE Healthcare). Similarly to the purification of HPP, the cells from 100 ml of the cell culture were resuspended in 10 ml of the resuspension buffer (50mM HEPES, 20mM NaCl, pH 7.4). The supernatant obtained after cell sonication was loaded on the 1-ml column equilibrated with resuspension buffer. Subsequently, the column was washed with 5 ml of washing buffer (50mM HEPES, 500mM NaCl, 0.5% Triton X-100, 150mM imidazol, pH 7.4) and AK was eluted with same washing buffer, however containg 250 mM imidazol. One-ml fractions were collected, checked on SDS-PAGE and fractions with sufficient purity were used for further purification step on gel filtration columns.

4.3.3 Protein concentration and storage

Protein samples were concentrated using Amicon Ultra Filters (Millipore). Basically two arrangements were used, depending on the volume of the same before and after concentration. Centricons were used for concentration using

centrifuge and concentration cells for concentration using ultrafiltration. Concentration was done at 4 °C according to manufacturer's recommendations.

Concentrated samples of HPP dimer were stored in 10-20 µl aliquots at -80 °C. Before subsequent crystallization, SAXS of mass-spectrometry experiments the sample was gently thawed on nice, centrifuged at 16000 g for 10-20 min and sample quality was checked on SDS-PAGE, blue-native electrophoresis, DLS and/or gel filtration. After 4 weeks the samples showed the same quality as freshly-prepared samples. Storage at higher temperatures proved to be not suitable due to generally low stability of α-HPP subunit [143].

4.3.4 Gel chromatography

The gel chromatography techniques were used to change the sample buffer, to check the oligomerization state of samples and to improve the purity of samples. Basically, two approaches were chosen – the simple gel filtration in the case of purification of AK and Fast Protein Liquid Chromatography (FPLC) using ÄKTA Purifier chromatographic system in the case of purification of HPP.

Concentrated sample of HPP was loaded on SuperoseTM 10/300 GL column (GE Healthcare) connected to FPLC system using 200 µl sample loop. The column was already equilibrated with buffer containing 25 mM HEPES, 20 mM NaCl, 5% glycerol, 2 mM $MnCl_2$, 0.1 mM DDM, pH 7.5 and the same buffer was used throughout the whole chromatographic run. Flow rate was 500 $\mu l.min^{-1}$ and 12 half-ml fractions were collected for the following SDS-PAGE analysis (**Figure 16**). The protein peaks were detected with a UV monitor at wavelength of 280 nm and the peak top was considered as the elution volume of given protein. BSA was used as the standard protein for column calibration.

In the case of AK, the sample was desalted on a Sephadex G-25 column (Sigma-Aldrich) using 20mM Tris-HCl, 150mM NaCl, pH 8.

4.3.5 Activity assay and protein quantification

Proteolytic activity of HPP towards AK as its substrate was evaluated as a change of protein mobility on SDS-PAGE ("gel shift assay"). Conditions for maximum activity as well as method as a whole were described previously [143]. Here we used this method as a tool for verification of native state of HPP dimer during various purification conditions.

Bradford assay was used for quantification of protein concentration. Absorbance was measured on micro spectrophotometer at wavelength of 560 nm.

4.3.6 Western Blot

Western Blot was routinely used to detect "His-tagged" proteins. In our case, the method uses SDS-PAGE gel electrophoresis to separate proteins by the length of the polypeptide chain. The proteins were then transferred to a nitrocellulose membrane, where they were stained with antibodies specific to the His-tag.

Following SDS-PAGE, the proteins were transferred to a nitrocellulose membrane (Biorad) in a transfer buffer containing 15 mM Tris, 192 mM glycine, 20% methanol, pH 8.3. The transfer was realized at 100 V for 1 hour in ice-cooled transfer cell. Subsequently, the membrane was blocked by 1% blocking solution (Blocking Reagent, Roche, dissolved in buffer containing 100 mM maleic acid, 150 mM NaCl, pH 7.5) and washed several times in PBS buffer with 0.05% Tween20.

For His-tag detection was used anti-His-tag antibody conjugated with horseradish peroxidase (Monoclonal Anti-polyHistidin Peroxidase Conjugate, Sigma, dilution ratio 1:2000). Immobilon Western Chemiluminiscent HRP was used as a substrate for subsequent chemiluminiscent detection of peroxidase activity on photographic paper (Kodak).

4.3.7 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a thermal-denaturation assay that measures the thermal stability of a target protein as a function of increasing

fluorescence caused by nonspecific binding of fluorescence dye (Sypro Orange, Sigma) to protein hydrophobic surfaces. In aqueous condition, the dye is strongly quenched by water. A sample containing native protein, whose hydrophobic surfaces are buried inside the protein core will not activate the dye. As protein gets denatured, hydrophobic surfaces of the protein gets exposed in the solution activating the fluorescence dye. Protein stability curve can be obtained by changing the temperature gradually to unfold the protein and measure the change in fluorescence in the time.

DSF was used to optimize purification and storage conditions for α-, β-subunit alone and for HPP dimer. Experimental samples were mixed in 96-well micro titration plate for real-time PCR as follows:

- 1 µl of protein sample (0.2 $mg.ml^{-1}$)
- 1 µl of Sypro Orange dye (40x diluted)
- 23 µl of buffer

Mixtures were kept on ice for 5 minute before subsequent fluorescence measurement on LighCycler 480 Multiwell (Roche) using HEX canal. Temperature gradient was from 20 to 80 °C, with 0.5 °C step, and length of equilibration ("dwell time") was 6 sec. In total, 48 different buffers were tested (**Table 3**).

	Buffer Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
	pH	**6.2**	**6.5**	**6.7**	**7.0**	**7.2**	**7.4**	**7.6**	**7.8**	**8.0**	**8.4**	**8.8**	**9.2**
Set A	Bis-Tris	20		20									
	NaCl	20	20	20	20	20	20	20	20	20	20	20	20
Set B	Bis-Tris Propane		20		20							20	20
	NaCl	200	200	200	200	200	200	200	200	200	200	200	200
Set C	Tris							20		20	20		
	NaCl	20	20	20	20	20	20	20	20	20	20	20	20
	Glycerol	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
Set D	Hepes					50	50		50				
	NaCl	200	200	200	200	200	200	200	200	200	200	200	200
	Glycerol	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%

Table 3. Buffers composition for DSF, divided in 4 groups (A-D). Units (mM) are omitted due to the limited space in table cells.

4.4 Construction of the homology model of HPP

Since the crystal structure of HPP is not yet available, we took the advantage of the crystal structure of homologous MPP (PDB ID: 1HR6) and built a homology model of HPP. For models construction, sequences of HPP subunits were aligned to MPP subunits using the ClustalX software (**Table 4**) [144]. A corresponding model structure was calculated by alignment mode at SWISS-MODEL server [145].

	α-subunits	**β-subunits**
Identity	16.9%	24.6%
Similarity	26.3%	43.2%

Table 4. Identity and similarity between HPP and MPP subunits. Calculated by ClustalW.

The final homology structure of HPP was obtained at the end of 20-ns-long non-restrained MD simulation, performed on the initial model to relax all possible strains that might have arisen during model construction. The protocol for this MD simulation is described in chapter dedicated to *Non-restrained MD simulations*. The structure stability was monitored during the whole production phase of MD simulation using the analysis of time-based and residue-based RMSDs and the analysis of secondary structure elements. The homology model showed stable conformation with the exception of the loops exposed to surface of the enzyme.

4.5 Protein crystallography

Protein crystallization experiments were performed in cooperation with the group of Dr. Řezáčová at Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

4.5.1 Dynamic light scattering

Dynamic light scattering (DLS) is a technique that can be used to determine the size distribution profile of small particles in suspension or polymers in solution. In

other words, it can be used to evaluate the homogeneity of any biological sample, especially of protein solutions for subsequent crystallization experiments.

The sample concentration for DLS experiment was ranging between 5 – 10 $mg.ml^{-1}$. Frozen sample was gently thawed on ice, centrifuged at 16000 g and 5 °C for 10 min. Subsequently, the sample was loaded into cuvette and experiment was employed at 20 °C. Number of DLS measurements in one experimental run was between 50 and 60. For data evaluation was used automatic built-in software.

4.5.2 Screening for crystallization conditions

For crystallization screening was used Crystal Phoenix crystallography dispenser with EasyXtal microplates. Experiments were performed with various HPP samples, differing in buffer composition and sample storage. "Sitting drop" method was used and ratio between protein sample and crystallization was 1:2 and 2:1. If the protein sample contained glycerol or DDM (or both) the same amount of the compound was added to crystallization screen. The crystallization temperature was 18 °C and following screens have been tested:

- JCSG+ Suite (Qiagene)
- PEGs Suite (Qiagene)
- MPD Suite (Qiagene)
- Hampton Index HT (Hampton Research)
- Structure Screen 1 (Molecular Dimensions)
- Structure Screen 2 (Molecular Dimensions)

4.6 Biological small-angle x-ray scattering (bio-SAXS)

All bio-SAXS experiments were performed in European Synchrotron Radiation Facility (ESRF) at beam line BM29 in Grenoble, France (application IDs: MX-1441 and MX-1491).

4.6.1 Samples preparation

SAXS data were collected from the wild-type and mutant form of HPP and MPP in the absence and presence of the peptide substrate. Protein samples were prepared according to the procedure described in *Protein expression and purification* and were brought to ESRF on dry ice. Subsequently, samples were gently thawed on ice and spinned down to remove possible protein aggregates. The protein concentrations were measured on NanoDrop and samples were diluted to the working concentration.

4.6.2 Experimental setup

Samples were put into quartz capilaries and measured using standard bio-SAXS setup of beamline BM29. SAXS data were collected at the bioSAXS beamline BM29 with a Pilatus 1M detector (Dectris Ltd., Baden, Switzerland) at a wavelength of 0.931 Å and a camera length of 2.42 m. Experimental hutch was further equipped with a marble table housing the modular-length flight tube and a sample handing equipment (automated sample changer with temperature control system). Measurements were performed at protein concentrations between 2 and 5 mg/ml to verify whether any interparticle effects that may have been present could be accounted for and rule out their influence on the analysis. Buffers controls were collected before and after data collection to provide an accurate solvent correction. To exclude the possibility of radiation damage, 10 frames, each of a 10-sec duration, were collected while continuously exposing fresh sample to the beam. The resulting frames were then compared to ensure no differences in the SAXS profiles were induced by exposure to x-rays.

4.6.3 Data processing

Row data collection, processing and preliminary analysis was performed in an automated manner using BM29 beamline dedicated software BsxCuBE. In the next steps, all SAXS data were processed using the ATSAS program package [146]. Radius of gyration (R_g) and maximum diameter (D_{max}) were evaluated from Guinier plots using PRIMUS software package and pair distance distribution functions, P(r), were computed with GNOM [147]. The solution shapes of samples were reconstructed from the high quality experimental data (GNOM functions) using the *ab initio* method and processed using Situs [148]. For each sample, 10 independent DAMMIF reconstructions were aligned, averaged, and filtered using the program package DAMAVER. The program CRYSOL was used to calculate scattering profiles based on atomic coordinates.

4.7 Hydrogen-deuterium (H/D) exchange

H/D experiments were performed in cooperation with the group of Dr. Novak at Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic.

4.7.1 H/D exchange experiment

The H/D exchange was initiated by a 10-fold dilution into a deuterated buffer (25 mM HEPES, 20 mM NaCl, 10% glycerol, pD 7.5). The molar ratio between the peptidase and peptide substrate was 1:10 and the protein concentration during H/D exchange was 0.3 $mg.ml^{-1}$. Sample with peptidase was incubated with its substrate for 10 seconds prior to the H/D exchange. Aliquots of 90 µl were taken after 30 s, 1 min, 3 min, 10 min, 30 min, 1 h, 3 h and 5 h. The exchange reaction was performed in ice bath with temperature ~4 °C and the reaction was stopped by quenching by the addition of 10 µl of 350 mM HCl followed by rapid freezing in liquid nitrogen. The experimental temperature 4 °C was chosen since at this temperature the peptidase is more stable and the proteolytic and the wild-type

HPP has only residual proteolytic activity (see the plots of substrate cleavage assay in **Figure 27** and **Figure 26**).

4.7.2 Digestion and HPLC separation

Each sample for the local kinetics was quickly thawed and injected onto an immobilized pepsin column (70 µL column volume, flow rate 100 µL/min, 0.4% formic acid in water, digestion time 40 sec). Peptides were trapped and desalted online on a peptide microtrap (Michrom Bioresources, Auburn, CA) for 1 minute at a flow rate 100 µL/min. Next, the peptides were eluted onto an analytical column (Halo C18, 0.3 × 50mm, 2.7 µm, 90 Å, Michrom Bioresources, Auburn, CA) and separated by a linear gradient elution 10-25% B in one minute, followed by 5 minute isocratic elution at 40% B at a flow rate 9 µL/min. Solvents were: A – 0.4% formic acid in water, B – 95% acetonitrile/0.4% formic acid.

In all analyses, injection and switching valves, immobilized pepsin column, trap cartridge and the analytical column were kept at ~4 °C (immersed into an ice-water bath) to minimize back-exchange. Outlet of the LC system was interfaced to ESI source of a mass spectrometer.

4.7.3 Mass spectrometry and data analysis

Mass spectrometric analysis was done on ESI-FT-ICR MS (9.4T Apex-Qe, Bruker Daltonics). For peptide mapping (HPLC-MS/MS) the instrument was operated in a data-dependent mode. Each MS scan was followed by MS/MS scans of the top three most intense ions. Tandem mass spectra were searched using MASCOT against a database containing sequence of the wild-type of mutant form of HPP. Sequence coverage was visualized using Draw Map script [149]. Analysis of deuterated peptides was done in HPLC-MS mode and the data were processed in DataAnalysis 4.0. Subsequently, the local kinetics data were processed by ExPro script provided by Gary H. Kruppa (personal communication) and transferred to DataAnalysis. Deuteration percentages were calculated as follows: % D = $100*((M_{pd} – M_{nd})/N)$ where M_{pd} is mass of partially deuterated peptide/protein,

M_{nd} is mass of non-deuterated and N is total number of exchangeable amide hydrogen within the peptide/protein. No correction for back-exchange was done. The figures showing deuteration kinetics for few selected peptides are shown in *Results* section and for all peptides used in our study are included in the *Supplementary material*.

4.7.4 MALDI-TOF substrate cleavage assay

MALDI-TOF MS analysis was performed to check cleavage efficiency of wild-type and E56Q mutant form of HPP incubated with peptide substrate derived from adenylate-kinase signal presequence. Two temperatures were chosen (4 °C and 37 °C) and the cleave reaction was stopped after 30 s, 1 min, 3 min, 10 min, 30 min, 1 h, 3 h and after overnight incubation.

4.8 Protein cross-linking techniques

Cross-linking experiments were performed in cooperation with the group of Dr. Novak at Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic.

4.8.1 Cross-linking experiment

Wild-type and E56Q mutant form of HPP proteins were cross-linked at a concentration of 1 $mg.ml^{-1}$ (25mM HEPES, 20mM NaCl, 5% glycerol, pH 7.5) with the homobifunctional cross-linkers disuccinimidyl suberate (DSS; Pierce) and disuccinimidyl glutarate (DSG; Pierce)[8]. Samples were incubated in a water bath at 10 °C for 5 min and the peptide substrate (derived from AK signal presequence - MLSTLAKRFASGKKDRM) was added to both samples at molar ratio 1:5 of enzyme to substrate. Subsequently, freshly prepared cross-linking reagents (DSG/DSGD4 and DSS/DSSD4) were added to the samples in 100X molar excess and cross-linking reaction mixtures were incubated in water bath at 10 °C for 30

[8] DSS and DSG both react with primary amines.

min. The reaction in the absence of cross-linking reagent was also carried out as a control. The cross-linking reaction was brought to a halt by 100 mM DTT present in the LDS NuPage electrophoretic sample buffer prior to the separation on SDS–PAGE (NuPage Bis-Tris 4-12% gradient gel). NuPage MES SDS buffer was used as a running buffer and Coomassie Briliant Blue R250 was used for gel staining. The bands of cross-linked protein were excised from the gel. The disulfide bonds were reduced with 50 mM TCEP for 5 min at 90°C and free cysteines were modified with 50 mM iodoacetamide for 30 min at 25°C in dark.

4.8.2 Sample digestion and LC-MS analysis

In gel proteolysis by trypsin (Promega) endoproteinase was carried out overnight at 37 ^{o}C with an enzyme/protein ratio of 1:20 (wt/wt). The peptide mixtures were desalted on a peptide MicroTrap column (Michrom Bioresources) prior to LC–MS analysis. After desalting, the peptide mixtures were loaded onto a reverse phased column MAGIC C18 column (0.2 × 150 mm, Michrom Bioresources) and separated on a capillary HPLC system (Agilent Technologies) at a flow rate of 4 µl.min^{-1} under the following gradient conditions: 1-10% B in 1 min, 10-45% B in 19 min, 45-95% B in 5 min, where solvent A was 0.2% formic acid, 2.5% acetonitrile and 2.5% isopropanol in water and solvent B was 0.16% formic acid in 90% acetonitrile and 5% isopropanol. The column was connected directly to an Apex-ULTRA Qe FT-ICR mass spectrometer (Bruker Daltonics) equipped with a 9.4 T superconducting magnet using an electrospray ion source. The instrument was calibrated externally using arginine clusters resulting in mass accuracy below 2 ppm.

4.8.3 Data analysis

Data acquisition and data processing were performed using ApexControl 3.0.0 and DataAnalysis 4.0 (Bruker Daltonics), respectively. The cross-links were identified using Links software [150]. The Links algorithm was set to consider the carbamidomethylation of cysteine and the possible single oxidation of methionine.

The mass error threshold was kept below 2 ppm and all assigned fragments were verified manually.

4.9 Molecular dynamics simulations

Molecular dynamics simulations were performed on the server operated by the group of Dr. Otyepka at Department of Physical Chemistry, Palacky University Olomouc.

4.9.1 Models for substrate translocation study

4.9.1.1 The GRL-bound structure

The *GRL-bound structure* of MPP in its initial interaction with a peptide derived from the malate dehydrogenase presequence (MDH: residues L^{2}SRVAKRA↓$FSST^{13}$; the arginine residue in position -2 relative to the cleavage site, i.e. the R-2 motif, is underlined) was built based on the MPP model described in Dvorakova-Hola et al. [118]. Here, the model of MPP with aldehyde dehydrogenase (ALDH) presequence peptide bound to GRL was constructed on the assumption that GRL may recognize its substrate in a similar way as the mitochondrial Tom20 receptor (part of translocase system of the outer mitochondrial membrane). The MPP structure with GRL-bound ALDH was generated by superposition with the analogous Tom20 receptor protein (PDB ID: 1OM2) and the model obtained was confirmed by tryptophan fluorescence experiments. In our study, the amino acid residues of the ALDH presequence were substituted by those of the MDH presequence and a final *GRL-bound structure* (**Figure 32**) was obtained at the end of a 100-ns-long non-restrained MD simulation performed on the initial model to relax all possible strains that may have arisen during model building.

4.9.1.2 The AS-bound structure

The *AS-bound structure* could not be built directly based on the crystal structure of MPP bound to MDH (PDB ID: 1HR9 with resolution 3.01 Å), since the structure of

the GRL misses here. Instead, the initial *AS-bound structure* was built based on the crystal structure of MPP bound to a peptide derived from the COX IV presequence (PDB ID: 1HR8 with resolution 2.70 Å) where the amino acid residues of the COX IV presequence were substituted by those of the MDH presequence. In the next step, the presequence was extended by building an additional four residues, FSST, at the C-terminus. The final *AS-bound structure* was obtained from a 100-ns-long non-restrained MD simulation that was carried out on the initial model to relax all possible strains that may have arisen during model building (**Figure 32**).

4.9.2 Models for the study of the structural role of GRL

Two MPP crystal structures deposited in Protein Data Bank were used for models preparation – the structure of MPP without and with bound peptide substrate in its active site (PDB ID: 1HR6 with resolution 2.50 Å and 1HR8 with resolution 2.70 Å, respectively). The structure of wild-type MPP (WT MPP) without bound substrate was taken directly from the 1HR6 structure. However, the model of WT MPP with bound substrate could not be generated from the 1HR6 structure, as the GRL segment (288-292) shows only weak electron density and is unresolved in the crystal structure. Thus, the WT MPP with bound substrate was constructed by combining the 1HR6 and 1HR8 structures – i.e. both structures were aligned and while the structure of the WT MPP was taken from the 1HR6, the structure of the bound presequence was taken from the 1HR8. The bound substrate was a synthetic peptide derived from the CytC oxidase IV signal presequence (COX IV) with residues S^{7}IRFFKPAT<u>R</u>T^{17}↓ (the arginine residue in position -2 relative to the cleavage site, i.e. R-2 motif, is underlined). A model of mutant MPP with deletion of the GRL was constructed based on a sequence alignment of the WT MPP α-subunit and the *Rickettsia prowazekii* processing peptidase, a peptidase which lacks a large part of the GRL sequence (**Figure 12**). In addition, the previous experimental work was also kept in mind [102, 129]. Thus, residues 285-300 of the α-MPP subunit were deleted, the N and C-termini of the resulting gap were linked

and two models were prepared, a ΔGRL MPP with the COX IV peptide substrate bound in its active site and a ΔGRL MPP with no bound substrate.

```
α-MPP of Saccharomyces cerevisiae:  GGGGSFSAGGPGKGMYSRL
α-HPP of Trichomonas vaginalis:     GGGSEFSSEGLGSGFSSLL
RPP of Rickettsia prowazekii:       FGGG----------MSSRL
α-M16 of Sphingomonas sp.:          LGGGG----------FESR
```

Figure 12. Alignments of the GRL regions of MPP and MPP-like proteins from selected organisms. The region containing residues 285-300 of the α-MPP subunit that is missing in the ΔGRL MPP models is underlined.

4.9.3 Targeted MD simulations

Targeted MD (TMD) simulations were employed to study the process of substrate translocation from the site of its initial recognition (i.e. GRL) to the MPP active site.

TMD simulations were performed in AMBER [151] with an additional term to the energy function based on the mass-weighted RMSD of a set of atoms in the GRL-bound structure compared to the AS-bound structure. The starting RMSD of the substrate, calculated based on comparison of both structures, was linearly decreased to 0 Å within a restrain period and then kept at 0 Å during the rest of the TMD simulation. Three different sets of input parameters were tested, differing in the duration of the restrain periods and the total length of the TMD simulation (**Table 5**). To prevent rotation of the entire molecule, the center of mass and orientation of the protein was fixed. Coordinates were stored every 2 ps.

	Starting RMSD	Restrain duration	Total duration
TD1	24.9 Å	0.5 ns	1.0 ns
TD2	24.9 Å	1.0 ns	1.2 ns
TD3	24.9 Å	1.6 ns	1.8 ns

Table 5. Targeted MD simulation parameters.

4.9.4 Non-restrained MD simulations

Non-restrained MD simulations were employed to study the functional and structural roles of GRL: specifically, (i) to study in detail the GRL-substrate interaction in two selected moments along the substrate translocation trajectory and (ii) to address the role of GRL in tertiary and quaternary structure of the MPP.

All non-restrained MD simulations were carried out using the AMBER suite [151] with the *parm99SB* force field [152]. The simulation protocol used was as follows. First, the protonation states of all histidine residues were set to create an optimal H-bond network. Next, all remaining hydrogen atoms were added using the *Leap* program from the AMBER package. The structures were charge-neutralized by adding an appropriate number or Na^+ ions. To prevent rotation of the entire molecule, the center of mass and orientation of the protein were fixed. All systems were inserted in a rectangular water box filled by TIP3P water molecules; the layer of the water molecules was 9 Å thick. Each system was then minimized prior to the production phase of the MD run in the following way. The protein was frozen and the solvent molecules and counter ions were allowed to move during a 1000-step minimization process followed by a 10-ps-long MD run under NpT conditions (i.e. p=1 atm, T=298.15 K).The side chains were then relaxed by several sequential minimizations with decreasing force constants applied to the backbone atoms. After relaxation, the system was heated to 50 K for 20 ps and then up to 298.15 K for 90 ps. The particle-mesh Ewald method for treating electrostatic interaction was used. For the production phase, all simulations were run under periodic boundary conditions in the NpT ensemble at 298.15 K and at a constant pressure of 1 atm using a 2-fs time integration step. The SHAKE algorithm with a tolerance of 10^{-5} Å was applied to fix all bonds containing hydrogen atoms. A 9.0 Å cutoff was used to treat non-bonding interactions. Coordinates were stored every 10 ps. The total durations of the production phases, the total number residues, atoms, counter ions and water molecules of the studied systems are summarized in **Table 6**.

	Total duration	Residues in			Ions		Water molecules	Atoms in total
		α-MPP	β-MPP	peptide	Zn^{2+}	Na^{+}		
Substrate in the MPP active site								
WT MPP	100 ns	457	439	0	1	6	21547	78507
WT MPP + peptide	100 ns	457	439	11	1	3	21505	78576
ΔGRL MPP	100 ns	441	439	0	1	7	22311	80614
ΔGRL MPP + peptide	100 ns	441	439	11	1	4	22239	80590
GRL-substrate interaction								
after 0.48 ns of TMD	100 ns	457	439	12	1	3	21834	79563
after 0.84 ns of TMD	100 ns	457	439	12	1	3	21749	79308

Table 6. Durations of production phases, and the number of amino acid residues, atoms and water molecules in the systems studied.

4.9.5 Data analysis

Several indicators were chosen for monitoring trajectory stability and conformational changes, including analyses of RMSDs, radius of gyration, secondary structure elements and the evaluation of inter-residue distances. Trajectories were analyzed and visualized using VMD package [153].

With regard to RMSD, three interpretations were used – time-based, residue-based and 2D RMSD analysis. In the case of time-based analysis, RMSDs was calculated in 0.1 ns intervals during the whole production part of MD simulation, as a measure of difference between the starting and present structures. RMSDs were also calculated for every residue between the starting structure and the structure obtained at the end of the MD simulation. All RMSDs were calculated using only the backbone Cα atoms MPP, those of the substrate were not monitored. The MDTRA software package was used for all RMSD calculations [154]. The radius of gyration was calculated using AMBER [151] and the buried area per monomer upon dimerization was analyzed using the PDBePISA Server [155]. All figures showing protein structures were generated by PyMOL software [156].

5. RESULTS

5.1 HPP expression and purification

Several purification approaches were employed in order to obtain highly pure, equimolar (1:1 ratio of α- and β-subunit in HPP dimer), stable and active form of HPP for subsequent crystallization, SAXS and mass-spectroscopy experiments.

5.1.1 Co-expression and pET Duet expression

Co-expression experiments using two separate vectors for α- and β-subunit of HPP resulted in non-equimolar samples, with excess of more stable β-HPP subunit. Following purification on FPLC didn't improve sample quality.

pET Duet experiments for simultaneous expression of the two HPP subunits from one vector resulted in very low expression level under different expression conditions. Although anti-His-tag western blot showed the presence of β-HPP subunit, HPP dimer was hardly visible on SDS-PAGE. Thus, due to the low yield the pET Duet system was abandoned.

5.1.2 Metal affinity purification of WT and E56Q HPP dimer

Pull-down metal affinity purification using C-terminally "His-tagged" β-HPP subunit and wild-type α-subunit (both expressed separately) did not lead to the HPP dimer having equimolar composition of its subunits – β-subunit was always in excess. With regard to the planned experiments, the quality of such a sample was considered as insufficient.

In the modified approach, N-terminally "His-tagged" α-HPP was expressed in inclusion bodies and renaturation procedure failed [143]. However, C-terminally "His-tagged" α-HPP subunit was expressed in native form and subsequent pull-down metal affinity purification together with wild-type β-HPP subunit resulted in proper dimer, with equimolar compositions of its subunits (**Figure 13**).

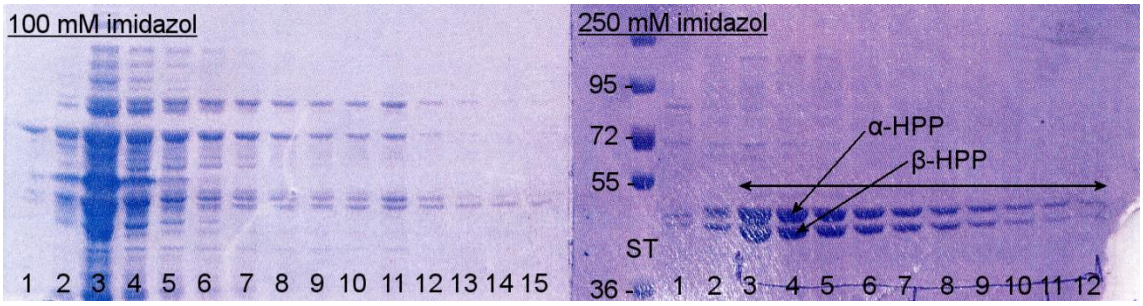


Figure 13. A typical example of metal affinity purification of HPP. On the left side are fractions eluted with buffer containing 100 ml imidazole. These fractions were not used for further processing due to a high amount of impurities (fractions 1–7) and low quality of HPP dimer (fractions 8-15). On the right side are fractions eluted with buffer containing 250 mM imidazole. The first two (or three) fractions were usually discarded and the rest of fractions (marked by two-side arrow) was combined and used for the next steps of purification. ST, protein standard with marked molecular sizes on its left side. 7% SDS-PAGE followed by Coomassie Brilliant Blue staining.

Washing buffers containing 100 mM imidazole but differing in glycerol concentration (5, 10 and 20%) were tested during the optimization of purification conditions. While the use of buffer with 5 or 10% of glycerol resulted in HPP dimer with non-equimolar compositions of its subunit (β-subunit was always in excess), the use of buffer containing 20% of glycerol led to the proper HPP dimer with equimolar compositions of α- and β-subunit.

Mutation E56Q was introduced to β-HPP active site to produce proteolytically inactive form of HPP. Inactive form of HPP was then purified as described above, using metal affinity purification and C-terminal His-tag of α-HPP subunit. Proteolytic activity of E56Q mutant of HPP was checked with adenylate-kinase (AK) as a substrate ("gel shift assay") and was shown to be less than 1% in comparison with the activity of WT HPP.

5.1.3 Purification of the mutant forms of α-HPP subunit

Tryptophan fluorescence spectroscopy was intended as a tool for characterization of the interaction of substrate with α-HPP subunit. For this purpose it was necessary to remove one native tryptophan residue in α-HPP subunit and introduce reporter one in the vicinity of α-HPP subunit GRL (**Figure 15**).

First, native W236 was mutated to tyrosine or phenylalanine (W236Y and W236F, respectively). Due to the rather low stability of α-HPP alone, the mutant forms of α-HPP (with His-tag at C-terminus) were purified in the form of dimer with WT β-HPP. As the purification tool was used metal affinity purification; conditions were already optimized during the purification process of WT HPP dimer.

Both W236Y and W236F mutants didn't seem to affect the association of α- with β-HPP subunit, since both HPP dimers was possible to purify using metal affinity purification (**Figure 14**). However, the level of expression of W236Y α-HPP subunit was much lower and led to reasonably lower purification yield (~30% in comparison with WT HPP).

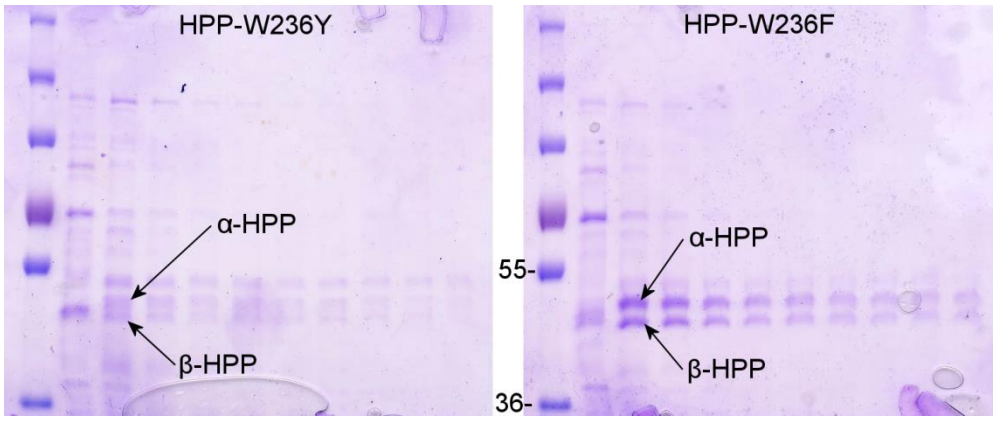


Figure 14. Metal affinity purification of W236Y and W236F α-HPP mutants in the form of HPP dimer. Protein standard in the first lines is followed by 10 one-ml fractions eluted from the column by 250 mM imidazole. 7% SDS-PAGE followed by Coomassie Brilliant Blue staining.

In the next step, both mutants were tested for proteolytic activity, using AK as the substrate ("gel shift assay"). While the activity of the HPP dimer with W236Y mutation α-subunit was below the limit of detection, the activity of W236F mutant was reduced "just" by half. Although both mutations in the α-HPP subunit didn't prevent it from association with β-HPP, they affected the enzyme activity of the HPP dimer as a whole. Thus, the W236F mutant was chosen for the subsequent introduction of the reporter tryptophan residues.

Based on the homology model of HPP four positions for reporter tryptophan residue were selected in the α-HPP subunit in the vicinity of its GRL (F256, F261, L267 and F271; **Figure 15**) and corresponding expression vectors were prepared. However, all these four mutations led to expression of the mutated α-HPP subunit exclusively in the inclusion bodies and not in the native form suitable for subsequent metal affinity purification. Thus, intended tryptophan fluorescence experiments could not be performed.

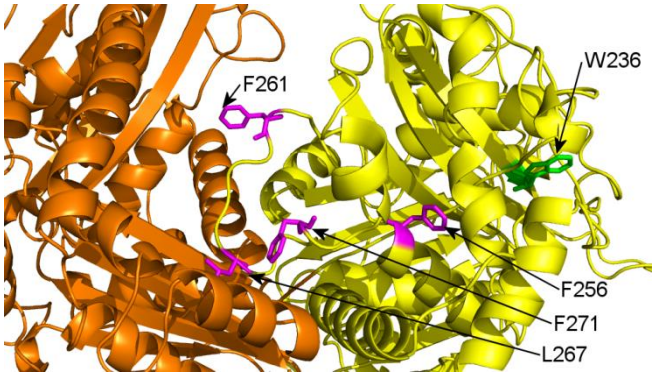


Figure 15. Homology model of HPP. α-HPP subunit in yellow and β-subunit in orange. Native W236 and four residues selected for introduction of the reported tryptophan residue are highlighted in green and magenta sticks, respectively.

5.1.4 Gel chromatography

Gel chromatography was used as a second purification steps of the HPP dimer. Following metal affinity purification, selected fraction were collected, concentrated, loaded on gel FPLC column (SuperoseTM 10/300 GL) and typically 12 half-ml fractions was collected (**Figure 16**). Selected fractions were again collected, concentrated and either straight away used for crystallization experiments and stored at -80 °C for further experiments (SAXS, H/D exchange, cross-linking).

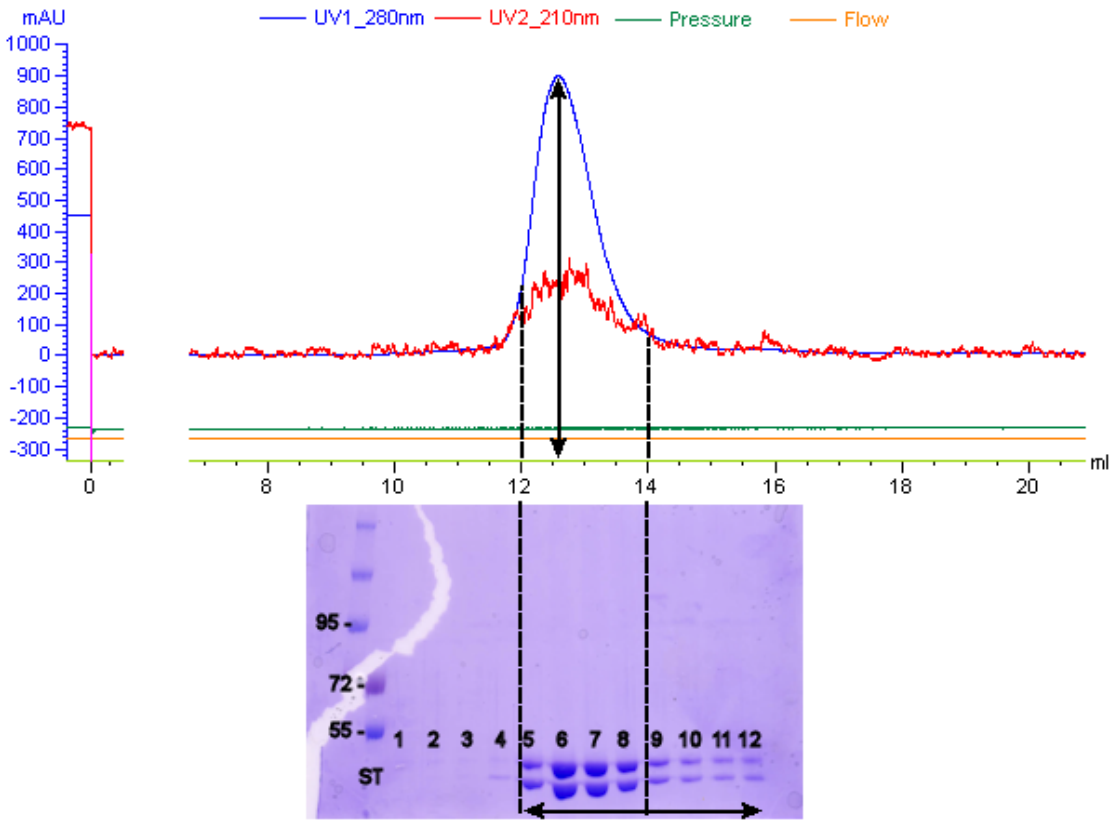


Figure 16. A typical example of gel chromatography of HPP dimer using ÄKTA Purifier system with attached SuperoseTM 10/300 GL column. ST, protein standard with marked molecular sizes on its left side. 7% SDS-PAGE followed by Comassie Brilliant Blue staining.

5.1.5 Storage conditions - DSF and DLS

Differential scanning fluorimetry (DSF) was used as a tool for initial optimization of storage conditions for α-, β-HPP subunit and for HPP dimer. In total, 48 buffers were tested (**Table 3** in *Material and Methods* section) differing in pH, NaCl and glycerol concentration. Although β-HPP subunit and HPP dimer are stable in relatively broad range of pH (6.2 – 8.4), the pH optimum for storage is between 7.0 – 8.0 (**Figure 17** and **Figure 18**, respectively). With regard to NaCl and glycerol concentration in the buffer both systems prefer 20 mM than 200 mM NaCl and 20% glycerol than none. The DSF results for α-HPP subunit reflect its rather general low stability and are not reliable (**Figure 19**).

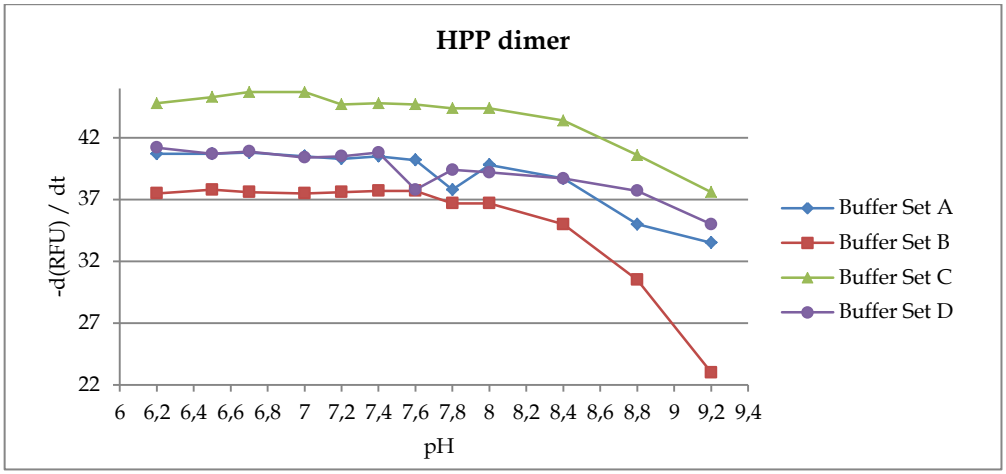


Figure 17. DSF experiment for HPP dimer. Buffers composition is summarized in Table 3.

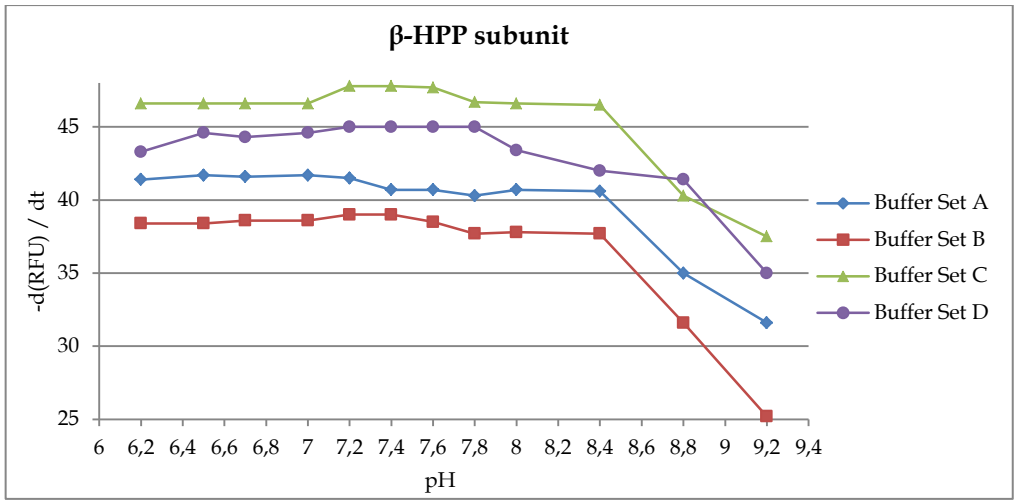


Figure 18. DSF experiment for β-HPP subunit. Buffers composition is summarized in Table 3.

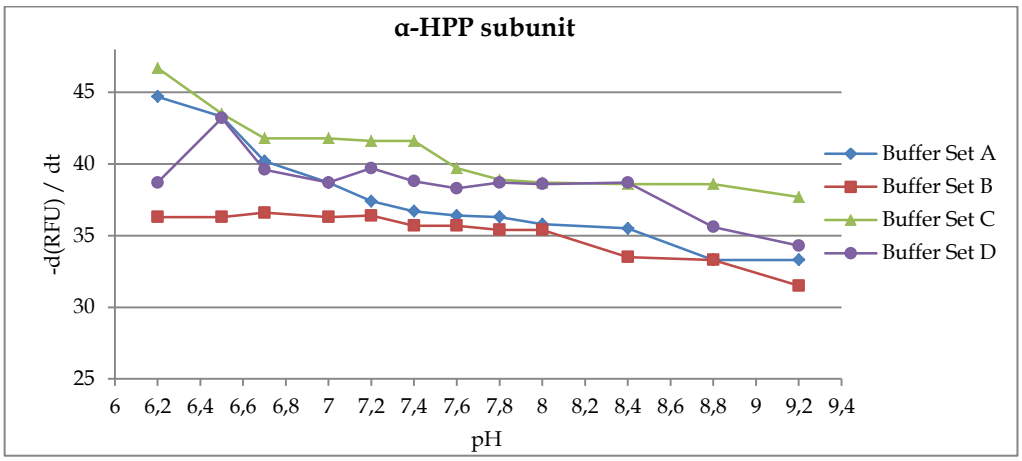


Figure 19. DSF experiment for α-HPP subunit. Buffers composition is summarized in Table 3.

Dynamic light scattering (DLS) was routinely used to verify sample homogeneity before crystallization experiments. Thawed samples appeared to have sufficient quality, even after 5 days of storage at -80 °C (**Figure 20**).

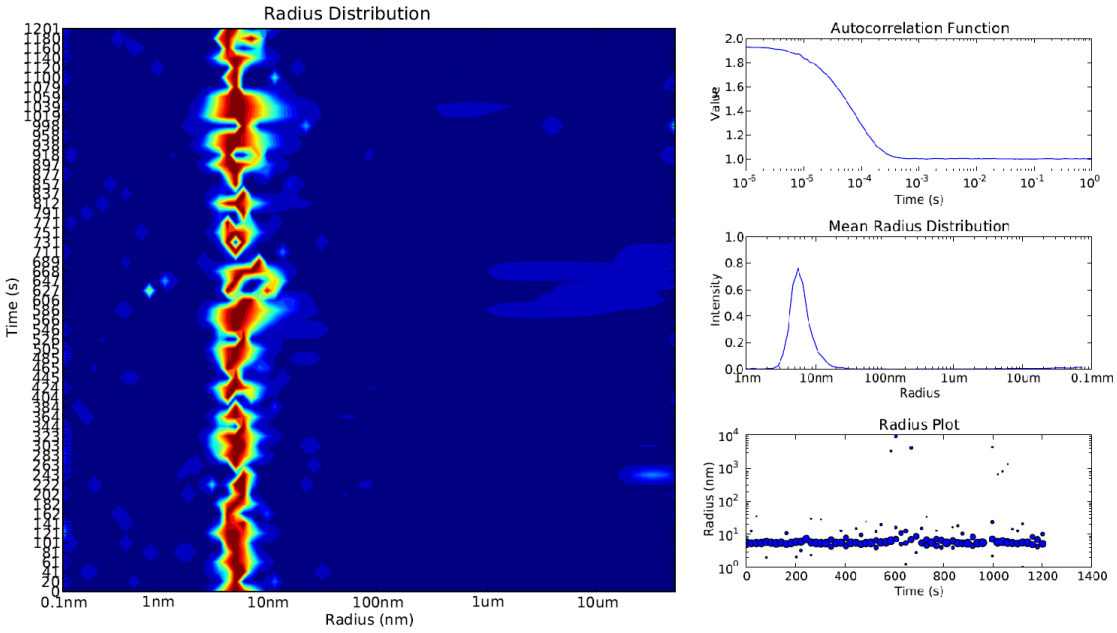


Figure 20. Example of DLS measurement of HPP dimer. The sample was stored ~12 hours in 25 mM HEPES, 20 mM NaCl, 5% glycerol, 2 mM $MnCl_2$, 0.1 mM DDM, pH 7.5 at -80 °C.

5.2 Crystallization experiments

All crystallization experiments were performed on WT HPP dimer and in total six crystallization screens were tested. Freshly prepared samples of HPP as well as samples stored at -80 C° were used and according to DLS there was no difference in sample quality (**Figure 20**). Screen compositions and general crystallization conditions were designed according to the previous work of Taylor et al. who solved the structure of homologous MPP dimer [88]. Thus, sitting-drop vapor diffusion method was used and crystallization experiments were performed at 21 °C. Sample concentration was ranging between 5 and 10 $mg.ml^{-1}$ and all protein samples contained DDM detergent.

MPP crystals were grown in the presence of PEG 10.000, ethylene-glycol and 2-methyl-2,4-pentanediol (MPD). Thus, apart from other screens, JCSG+,

PEGs and MPD Suite crystallization screens from Qiagen were used. Indeed, these three crystallization screens seemed to be more suitable for HPP crystallization since provided micro-crystallic or light precipitates, protein spherules or "stars" and in some cases even crystal needles (**Figure 21**).

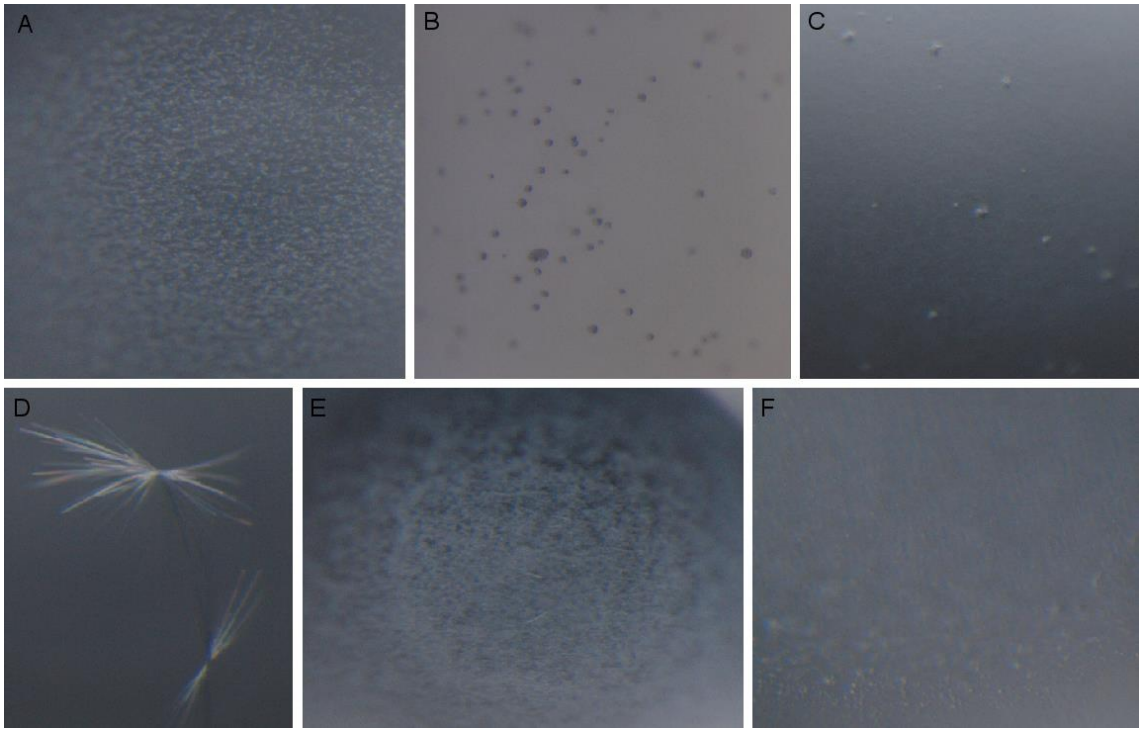


Figure 21. A few examples of crystallization experiments with WT HPP. A, E, F - Micro-crystallic precipitate. B – protein spherules. C, F – "stars". D – needles. For composition of screen buffers see **Table 7**.

Figure	Screen Name	Screen Buffer	Results
A	JCSG+ Qiagen	0.2 Sodium chloride, 0.1 M HEPES pH 7.5, 10% Isopropanol	Micro-crystallic precipitate
B	MPD Qiagen	0.1 M Imidazol-HCl pH 8.0, 35% MPD	Spherules
C	PEG Qiagen	0.1 MES pH 6.5, 40% PEG 200	"Stars"
D	MPD Qiagen	0.2 Magnesium acetate, 0.1 M MES pH 6.5, 15% MPD	Needles
E	JCSG+ Qiagen	0.2 M Lithium sulfate, 0.1 M Tris pH 8.5, 1.26 M Ammonium sulfate	Micro-crystallic precipitate
F	PEG Qiagen	0.1 M Sodium HEPES pH 7.5, 40% PEG 200	Micro-crystallic precipitate / "Stars"

Table 7. Composition of selected screen buffers. Column *Figure* refers to **Figure 7**.

Overall, crystals of WT HPP suitable for subsequent diffraction experiments have not been so far obtained. The system consisting of

proteolytically inactivate E56Q mutant mixed with peptide substrate derived from the adenylate-kinase are currently being performed.

5.3 Small-angle X-ray scattering

Bio-SAXS method was employed to study the effect of the presence of the peptide substrate in peptidase active site on the tertiary and quaternary structure of HPP and MPP. Both WT and proteolytically inactive forms were studied – the inactive forms of HPP and MPP were prepared by the introduction of E56Q and E73Q point mutations, respectively.

5.3.1 Hydrogenosomal processing peptidase

In the case of HPP, the changes in the shape and size of WT or mutant HPP were not observed, regardless the presence or absence of the peptide substrate in the sample (**Figure 22 –** A and B). However, an important difference between the scattering curves of the WT and mutant peptidase was found, both with and without peptide substrate present in the samples (**Figure 22 –** C). The two enzymes have roughly the same molecular weight (~70-75 kDa) since the zero intensities I(0) for the two enzymes are the same. However, both enzymes have different shapes because a discrepancy around s=1.2 Å was observed. The combination of the same molecular weight but different scattering profiles suggests different conformation between the WT and mutant form. The difference in scattering profiles is not due to sample aggregation since a potential aggregation would have increased the zero intensities (and the molecular weights) of the enzymes.

Theoretical scattering curves were calculated by Crysol for homology model of HPP, crystal structure of MPP and the two conformations of the *Sphingomonas sp.* M16B peptidase (SPH) – open and closed (**Figure 22 –** D). Both SPH theoretical scattering curves are very similar in shape, although an important difference exists, interestingly also around s=1.2 Å. However, their similarity with experimental scattering curve for HPP is very distant. On contrary, the theoretical

scattering curve of HPP comes closer to the experimental data and has similar trend.

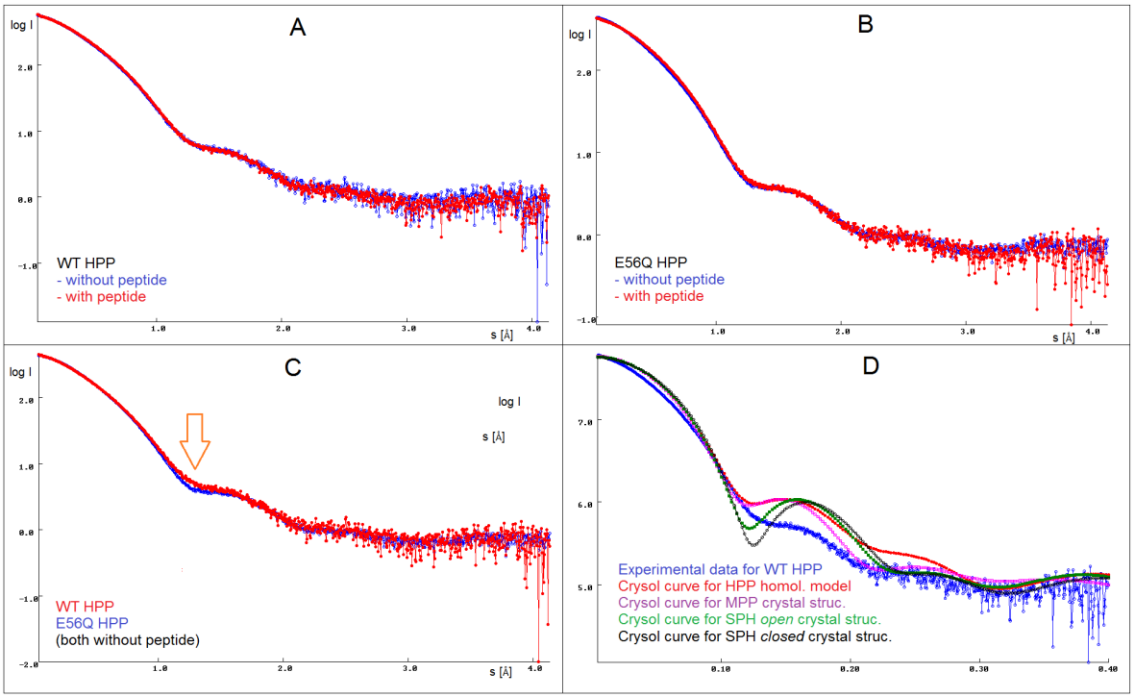


Figure 22. bio-SAXS scattering curves for WT and E56Q mutant form of HPP. The scattering curves on panels A and B show no conformation change in the shape and size of the WT peptidase neither E56Q mutant when a peptide substrate is added to the enzyme, respectively. Panel C shows the comparison of scattering curves for WT and E56Q mutant – note the discrepancy around s=1.2 Å denoting the difference in peptidase shapes and sizes (orange arrow). Panel D shows the experimental scattering curve for WT HPP in comparison with for theoretical scattering curves calculated by Crysol for homology model of HPP (red), crystal structure of MPP (pink) and for open and closed conformation of the *Sphingomonas sp.* M16B peptidase (SPH; green and black, respectively).

Further, bio-SAXS were used to compute theoretical solution-shape of WT and E56Q proteolytically inactive mutant form of HPP (**Figure 23**). The bio-SAXS models obtained are compacted and very similar in shape. Based on the crystal structure of homologous HPP we expected a globular shape of HPP and this expectation was confirmed. The comparison of the pair-distance distribution functions showed that the mutant HPP forms a more elongated shape that the WT - gyration radius of 42 Å versus 36 Å and maximum diameter of 140 Å versus 120 Å. The values calculated on the homology model using Crysol are 28 Å and 95 Å, respectively.

The "tail extension" of solution-shape models of HPP is remarkable and we can just speculate if this is a real structural trait of HPP or just artifact arisen from sample aggregation. The fact is that the "tail extension" was present in all solution-shape models (both WT and E56Q) reconstructed based on the experimental data acquired during three independent bio-SAXS experiments with various samples, differing in the purification process and storage conditions.

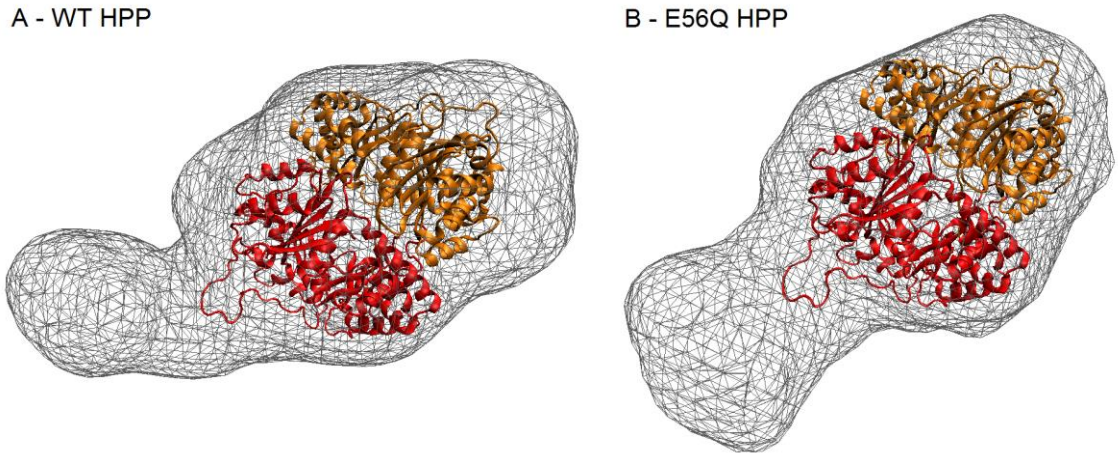


Figure 23. Representative examples of rigid-body docking of HPP homology model into bio-SAXS *ab-initio* solution-shape models of (A) WT HPP and (B) E56Q mutant. Volumetric contour (i.e. solution shape) map is visualized using VMD [153] and was created using Situs program package [148] as follows: (1) mass-weighting no, (2) B-factor threshold no, (3) voxel spacing 2 Å, (4) kernel width 6 Å, (5) Gaussian smoothing kernel, (6) lattice correction yes and (7) kernel amplitude scaling factor 1. The α-HPP and β-HPP subunit is displayed as a cartoon in red and orange color, respectively. Both homology models were structurally aligned to have the same orientation. The "tail extension" on the left side of both mesh representations is very likely artificial and was present in all evaluated models.

5.3.2 Mitochondrial processing peptidase

Similarly to the case of HPP, bio-SAXS method was used to study the effect of peptide substrate bound in peptidase active site on tertiary and quaternary structure of the MPP dimer. As it has been already mentioned, the MPP crystal structure was solved without and even with active-site-bound peptide substrate [88]. According to this study the MPP dimer is kept in partially open-closed conformation, regardless the presence or absence of substrate in its active site. Bio-SAXS experiments confirmed this idea – in the case of WT MPP there is no change

in the shape and size of enzyme dimer when a peptide substrate is added to the sample (**Figure 24**– Panel A). The scattering curves for E73Q proteolytically inactive MPP mutant could not be evaluated due to the sample aggregation before and/or during SAXS data collection, as is illustrated by mismatch in zero intensity for the sample with and without peptide substrate (**Figure 24 –** B). In this case, purification procedure and storage conditions must be further optimized.

Theoretical scattering curves were calculated for crystal structure of MPP, homology model of HPP and the two conformations of SPH – open and closed (**Figure 24 –** C). Since MPP experimental scattering curve is in good agreement with the theoretical one, these two profiles represent a kind of "method validation". While the scattering profile of HPP homology model follows also the experimental data to some extent, the theoretical scattering curves of SPH are quite distant.

Panel D in **Figure 24** shows the solution-shape contour of *ab initio* solved MPP structure with docked crystal structure of MPP. Similarly to the case of solution-shape model of HPP the MPP model has the same limitations, including the "tail extension" which might arose due to the sample aggregation. The gyration radius of calculated solution-shape model was 35 Å and maximum diameter 120 Å, while the corresponding parameters calculated by Crysol are 32 Å and 91 Å.

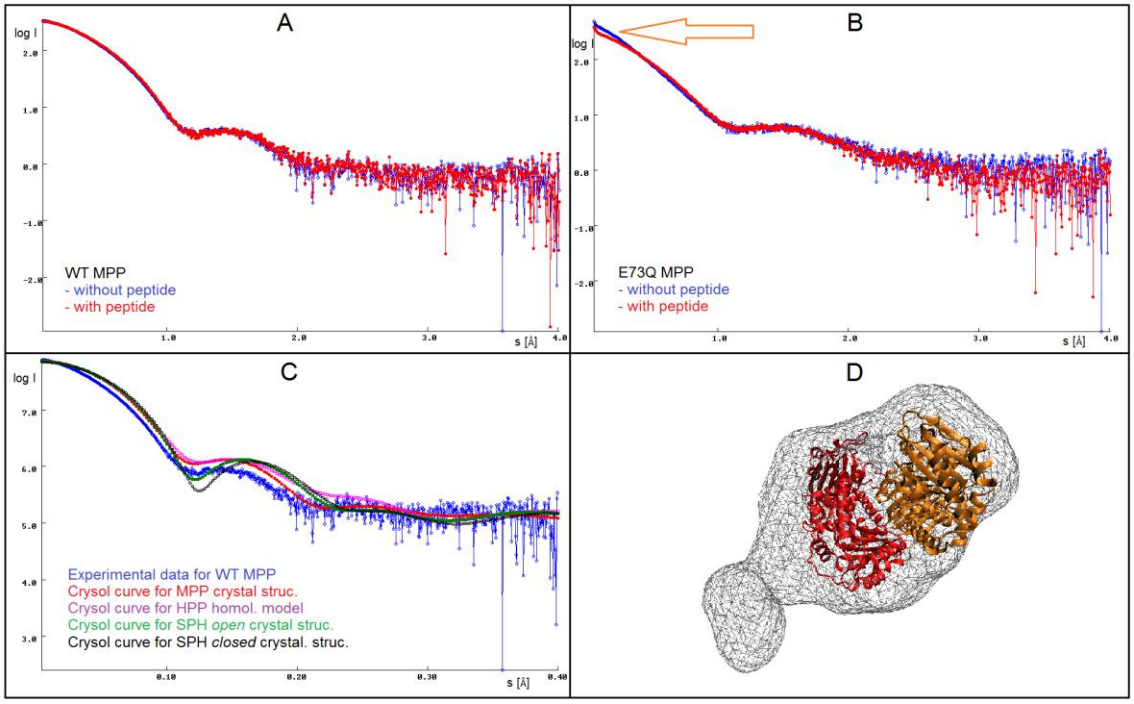


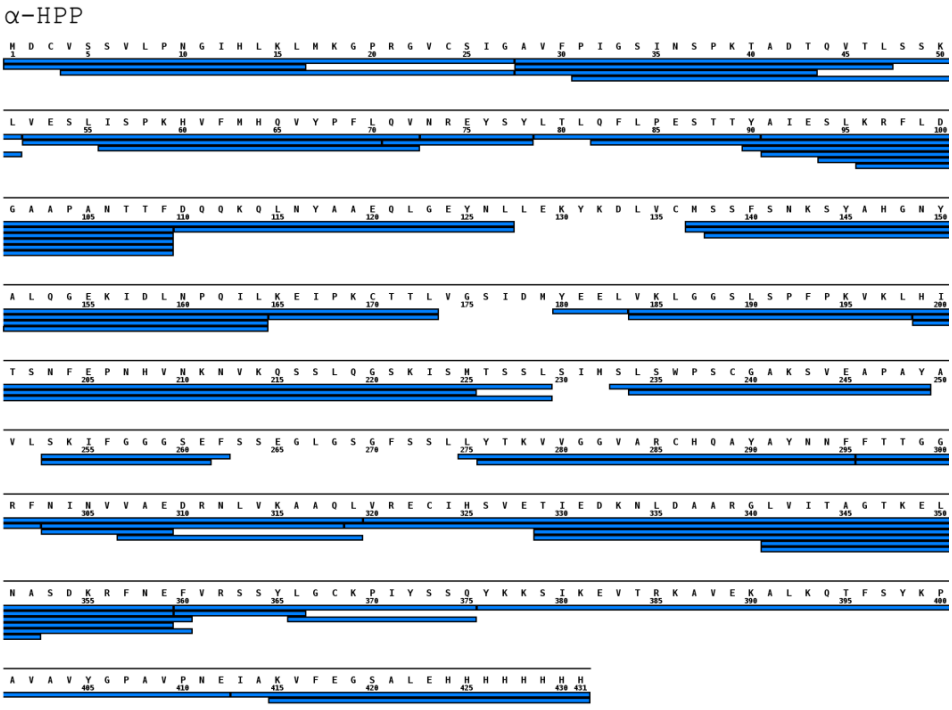
Figure 24. bio-SAXS scattering curves for WT and E73Q mutant form of MPP. The scattering curves on panel A show no conformation change in the shape and size of the WT peptidase when a peptide substrate is added to the enzyme. Panel B shows the same information for E73Q mutant – in this case, however, the SAXS data was not possible to interpret due to the aggregation of E73Q MPP sample before and/or during SAXS data collection (orange arrow). Panel C shows the experimental scattering curve for WT MPP in comparison with for theoretical scattering curves calculated by Crysol for crystal structure of MPP (red), homology model of HPP (pink) and for open and closed conformation of SPH (green and black, respectively). Note that MPP experimental and theoretical scattering curves have a very similar trend. Panel D shows representative example of rigid-body docking of MPP crystal structure into bio-SAXS *ab-initio* solution-shape models of WT MPP. The α-MPP and β-MPP subunit is displayed as a cartoon in red and orange color, respectively. For further visualization details see **Figure 23**. The "tail extension" on the left side of the mesh representation is very likely artificial and was present in all evaluated models.

5.4 Hydrogen-deuterium exchange

To identify the substrate binding region and to study the conformation changes during the process of the substrate binding in the HPP active site, the deuteration experiments were performed on the WT and E56Q mutant form of HPP both in the in the presence and the absence of the peptide substrate derived from the adenylate-kinase signal presequence (MLSTLAKRFASGKKDRM).

5.4.1 Peptide mapping

Wild-type HPP was digested on pepsin column as described in *Material and method* section and LC MSI-MS/MS analysis was performed. On the basis of the MS/MS spectra, 52 and 87 peptides suitable for H/D exchange analysis were identified, covering 92% and 97% of the α-HPP and β-HPP subunit sequence, respectively (**Figure 25**).



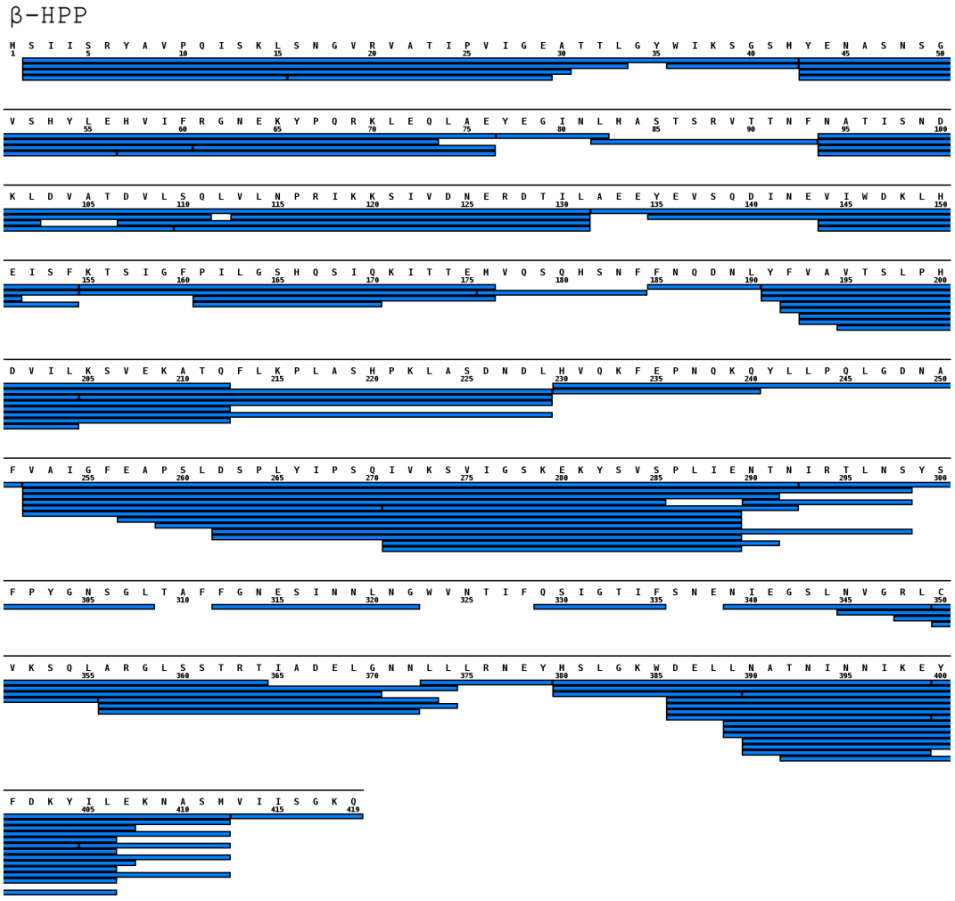


Figure 25. Peptide mapping of α- and β-HPP subunits after immobilized-pepsin digestion used in HD-MS experiments. All peptides were identified by ESI-Trap-MS/MS analyses and are represented under the amino acid primary sequence by blue bars.

5.4.2 MALDI-TOF substrate cleavage assay

The proteolytic activity of WT HPP and inactivity of E56Q HPP was also verified by a simple cleavage reaction followed by MALDI-TOF analysis. Enzyme was mixed with peptide substrate derived from adenylate-kinase signal presequence, reaction mixtures were kept in water bath at 4 or 37 °C and aliquots were taken after 30 sec, 1 min, 3 min, 10 min, 30 min, 1 hour, 3 hour and "overnight" of incubation for subsequent MALDI-TOF analysis.

The thermal optimum for HPP activity was shown to be 37 °C [143] and after 10 min 99% of peptide substrate was cleaved. On contrary, at 4 °C and following overnight incubation only ~50% of the substrate was cleaved (**Figure**

26). In the case of E56Q HPP mutant, less than 1% of substrate was processed following 1 hour and overnight incubation, at 37 °C and 4 °C, respectively.

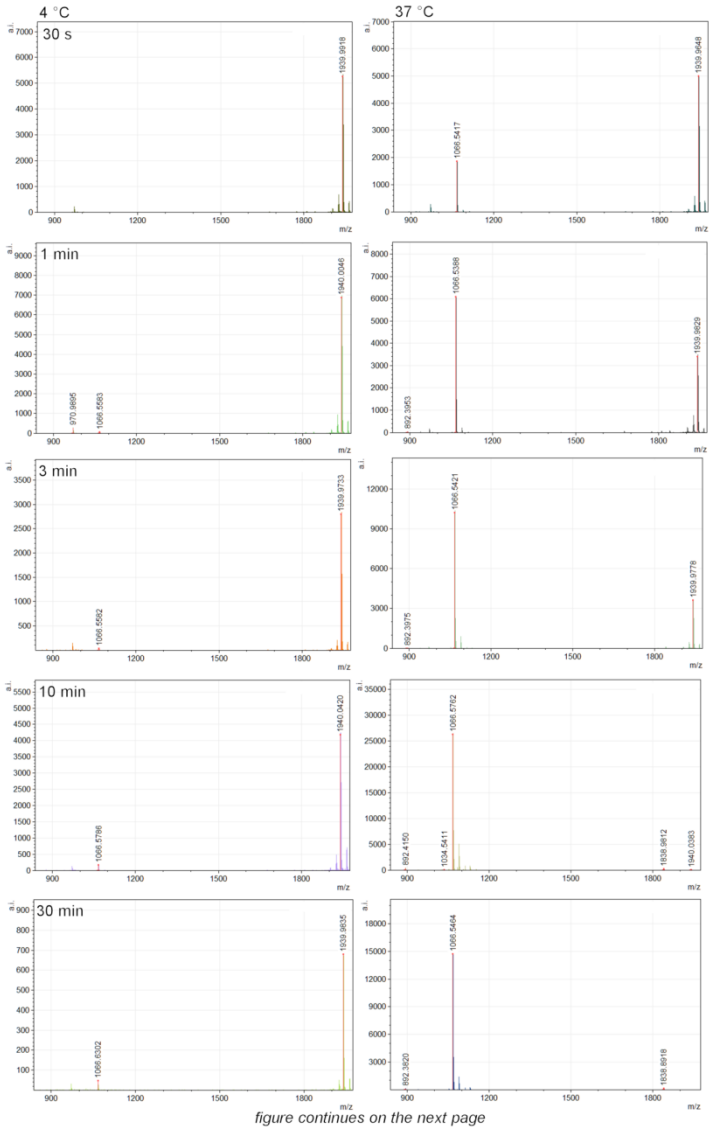


figure continues on the next page

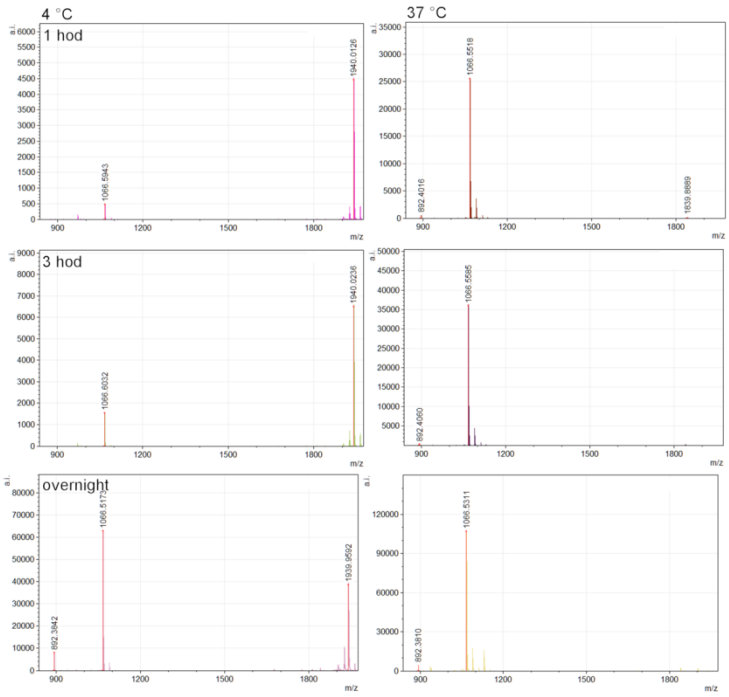


Figure 26. Substrate cleavage assay for WT HPP and peptide substrate derived from AK signal presequence (MLSTLAKRFASGKKDRM). The non-cleaved peptide substrate has molecular mass of 1940 Da. After cleavage, two peptides were obtained, having a molecular mass of 892 and 1066 Da.

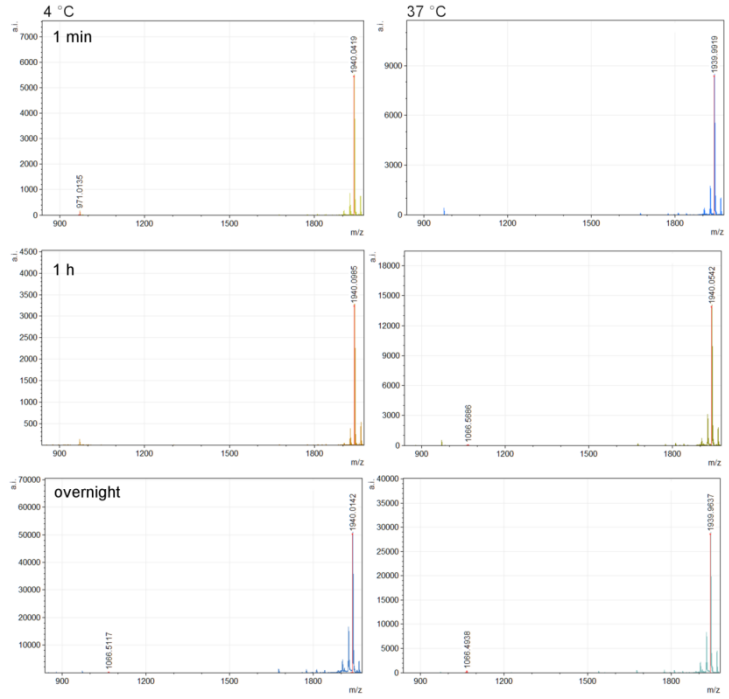


Figure 27. Substrate cleavage assay for E56Q mutant form of HPP. The non-cleaved peptide substrate (MLSTLAKRFASGKKDRM) has molecular mass of 1940 Da. After cleavage, two peptides were obtained, having molecular mass of 892 and 1066 Da. Only 3 time-steps are shown.

5.4.3 Kinetics of H/D exchange for HPP

We expected that peptide bound in HPP active site could shield some β-HPP residues in its neighborhood and thus would cause a decrease in the H/D exchange rate in this region. However, such a decrease in H/D exchange rate was not visible in kinetics curves and, thus, the binding of the peptide substrate in HPP active site was not possible to confirm (but neither exclude). Further, the kinetics curves suggest that HPP dimer does not undergo conformation changes depending on whether the peptide substrate is present in HPP active site or not (**Figure 28**).

Interestingly, from the kinetics curves is obvious that the single E56Q mutation of HPP active site does affect the conformation of whole HPP dimer. In the case of β-HPP subunit, the WT form of HPP dimer exhibits higher H/D exchange rates than E56Q mutant. Thus, the mutation seems to somehow "stabilize" the β-HPP subunit. On contrary, in the case of α-HPP subunit, the higher H/D exchange rates were observed for E56Q mutant form of HPP dimer.

The change of dimer conformation seems to be caused mainly by changes on the β-HPP subunit since the changes in deuteration kinetics of α-HPP are only minor. Specifically, only two α-subunit peptides with higher exchange rate were identified - peptide 360-366 and 366-375 (**Figure 28 –** right part). Both showed a shift in the deuteration level since one hour and more but no earlier. On contrary, the deuteration kinetics of β-HPP subunit was largely affected and a number of peptides representing the process was identified (**Figure 28 –** left part). In addition to the common EX2 exchange mechanism, a specific type EX1 was identified in some β-HPP peptides (for instance 329-335 and 271-289). While the EX2 exchange rate is continuous the EX1 is stepped and suggests that given part of protein remains in "locked conformation" for longer time.

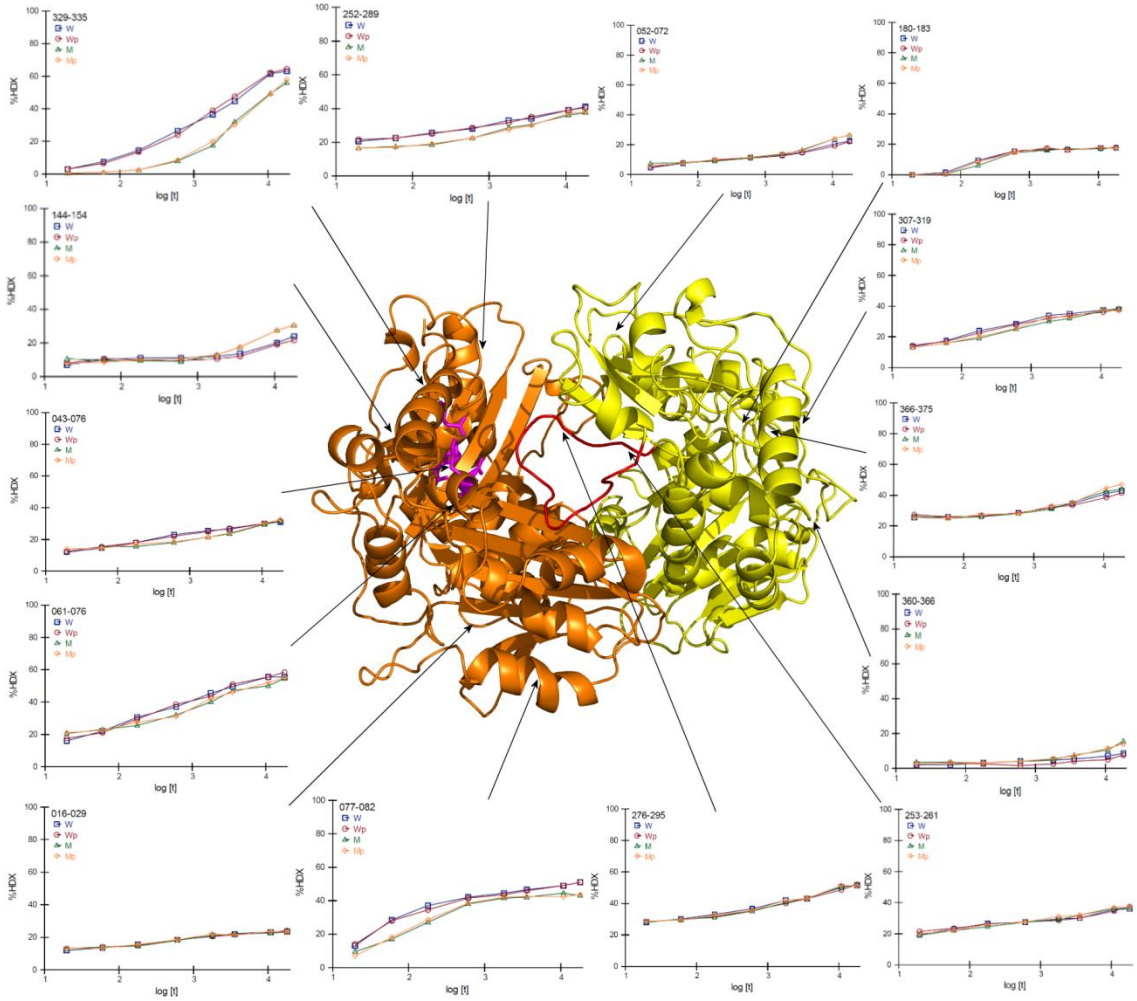


Figure 28. Deuteration kinetics for H/D exchange for selected peptides generated from WT and E56Q mutant form of HPP both with and without peptide substrate by pepsin proteolysis. Kinetics curves for peptides generated from WT without and with peptide are displayed in blue and red color, respectively, and the peptides from E56Q mutant are displayed in green and yellow color, respectively. Kinetics is shown for times between 30 sec and 5 h of H/D exchange (the x-axis corresponds to logarithmic scale of time in seconds). The y-axis represents percentage of H/D exchange. The position of the first and last amino acid of the peptide is shown on the top left corner of each plot. In the middle part of figure is displayed homology model of HPP with α- and β-HPP subunit colored in yellow and orange, respectively. The loop corresponding sequentially to the GRL of α-MPP is displayed in red and zinc-binding motif residues are displayed as pink sticks. Note that (i) there is no difference in H/D exchange rate when peptide substrate is added to the sample and that (ii) the H/D exchange rate for peptides derived from β-subunit is higher for the WT form, while for the peptides derived from α-subunit is higher for E56Q mutant form.

Further, the effect of zinc-ion on HPP dimer conformation was studied both on WT and E56Q mutant form of HPP. Since previous findings demonstrated that the peptide substrate does not affect HPP conformation, this type of H/D exchange experiments was performed only with samples without peptide

substrate. Basically, two sample buffers were used – one with $MnCl_2$[9] and second with EDTA. Kinetics curves suggest that the effect on Mn^{2+} ions is minor but exists. Similarly to E56Q proteolytically inactive mutant form of HPP the β-HPP subunit was affected more than α-subunit that seems to be affected only indirectly (**Figure 29**). Interestingly, the majority of β-HPP peptides generated from the samples with EDTA exhibited lower exchange rates (i.e. "stabilizing effect)", similarly to the effect of E56Q mutation. The trends in exchange rates were the same both for the WT and E56Q mutant form of HPP.

[9] Although HPP is zinc-peptidase the experiments were performed in the presence of Mn^{2+} ions. Previously we showed that HPP is active in broader range on Mn^{2+} concentration than Zn^{2+} [143]. In fact, zinc ions act as HPP inhibitor from a certain concentration.

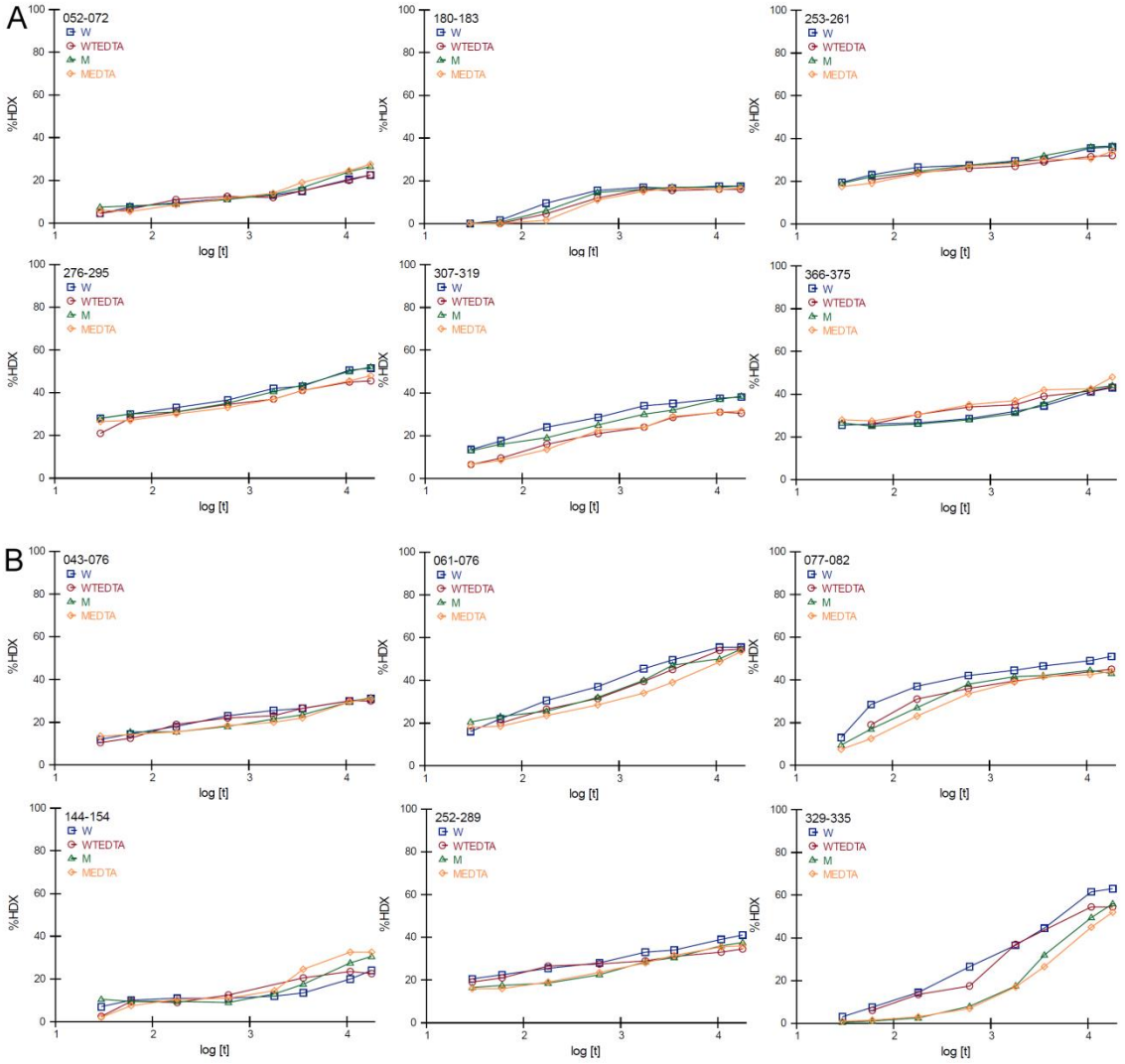


Figure 29. Deuteration kinetics for H/D exchange for selected peptides generated from WT and E56Q mutant form of HPP both in the absence and presence of EDTA by pepsin proteolysis. Kinetics curves for six selected peptides generated from α- and β-HPP subunit are displayed on the panel A and B, respectively, and the peptide selection is the same as on Figure 28. Kinetics curves for peptides generated from WT without and with EDTA are displayed in blue and red color, respectively, and the peptides from E56Q mutant are displayed in green and yellow color, respectively. Kinetics is shown for times between 30 sec and 5 h of H/D exchange (the x-axis corresponds to logarithmic scale of time in seconds). The y-axis represents percentage of H/D exchange. The position of the first and last amino acid of the peptide is shown on the top left corner of each plot.

Changes in the isotopic profiles of representative peptides generated by pepsin digestion of WT and E56Q mutant form of HPP as a function of H/D exchange time are illustrated in **Figure 30**.

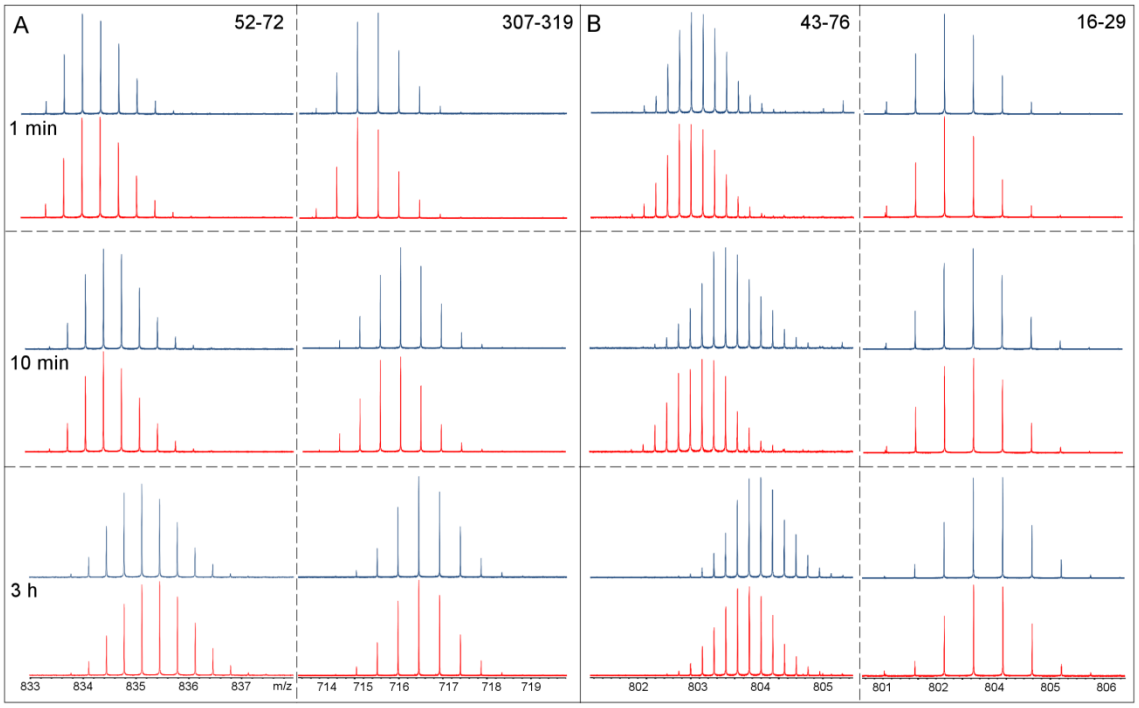


Figure 30. Detail of isotopic profiles of selected peptides. Panel A and B show isotopic profiles of two selected peptides generated by pepsin digestion of α- and β-HPP subunit, respectively. Blue and red plots correspond to WT and E56Q mutant form of HPP. The native sequence numbering is shown at the top of each column. The three rows show changes in the isotopic distribution of each peptide for three distinct time-points of the H/D experiment (1 min, 10 min and 3 hours).

5.5 Chemical cross-linking of HPP

The cross-linking method was employed (i) to describe inter and intra HPP subunit interactions, (ii) to verify the presence of peptide substrate in the HPP active site and (iii) to validate the homology model of HPP.

Protein sequence was completely covered and all cross-linked residues derived from the identified cross-linked peptides are listed in **Table 8**. For example, the signal at m/z 1869.0096 (error of 0.8 ppm) was assigned as a cross-link between lysines K348 and K355 in peptide G340-R356. Out of 47 cross-linked residues identified, 34 were formed with DSG and 13 with DSG. Out of 34 unique cross-linked peptides identified, 18 were assigned to α-HPP subunit, 8 to β-HPP and 10 to HPP dimer. Only one cross-linked peptide was assigned to the complex of α-HPP subunit and peptide substrate. On contrary, no cross-linked was assigned to the complex of β-HPP subunit and peptide substrate.

Agent	Cross-linked residues	Distance [Å]
α-HPP subunit		
DSG/DSS	K386-K390	5.9
DSG	K348-K355	10.1
DSG	K390-K393	5.4
DSG	K195-K197	6.3
DSG/DSS	K378-K381	5.1
DSG	K393-K399	7.4
DSG	K390-K399	10.3
DSG	K386-K393	10.4
DSG	K132-K386	30.7
DSG	K132-K390	23.4
DSG	K165-K195	12.3
DSG	M1-K214	32.5
DSG	K169-K355	23.6
DSG	K156-K214	22.6
DSG	K39-K355	32.4
DSG	K18-K355	12.4
DSG	K15-K348	21.9
DSG	K15-K355	14.5

Agent	Cross-linked residues	Distance [Å]
β-HPP subunit		
DSG/DSS	K119-K120	3.8
DSG	K279-K281	6.1
DSG	K403-K408	7.8
DSG/DSS	K65-K209	15.9
DSG/DSS	K65-K215	20.1
DSS	K65-K279	50.4
DSS	K14-K352	29.9
DSS	M1-K352	40.9

Agent	Cross-linked residues	Distance [Å]
HPP dimer		
DSG	K390-K418	66.4
DSG	K393-K418	65.1
DSG/DSS	K355-K281	32.5
DSG	K378-K398	57.9
DSG/DSS	K377-K120	44.5
DSG/DSS	K378-K120	46.2
DSG/DSS	K381-K120	42.4
DSG/DSS	K59-K279	30.8
DSG	K381-K101	15.5
DSG	K333-K408	62.9
α-HPP subunit and peptide		
DSG	K14-K197	-

Table 8. The list of all identified peptides cross-linked either with DSG, DSS, or both. Table is divided into 4 parts – the peptides identified in α- and β-HPP subunit, the HPP dimer and in a complex of α-subunit with peptide substrate. The column *Distance* shows the distance between Cα atoms of cross-linked residues measured on HPP homology model.

The distances between Cα atoms of cross-linked residues calculated based on the homology model of HPP are summarized in **Table 8**. Since the spacer arm of DSS and DSG is 11.4 Å and 7.7 Å, respectively, following cutoffs were used: 22 Å and 20 Å between Cα atoms of lysine residues cross-linked with DSS and DSG, respectively. Thus, the comparison of calculated distances with cutoff values provides insights into the validity of the homology model of HPP.

Considering the cutoff value for DSG and DSS, 11 out of 18 and 5 out 8 identified cross-links for α- and β-HPP subunit fulfill this criterion and thus the model of β-HPP subunit seems to have higher validity than α-subunit, although still very limited (**Figure 31**). On contrary, in the case of HPP dimer only one distance between lysine Cα atoms out of 10 is below the cutoff value.

Interestingly, the K65-K279 intramolecular cross-link in β-HPP subunits suggests that β-subunit might be capable of forming "head-to-tail" homodimers since K65 and K297 residues are located on the opposite sites of the homology model of HPP and the distance between them is 50.4 Å.

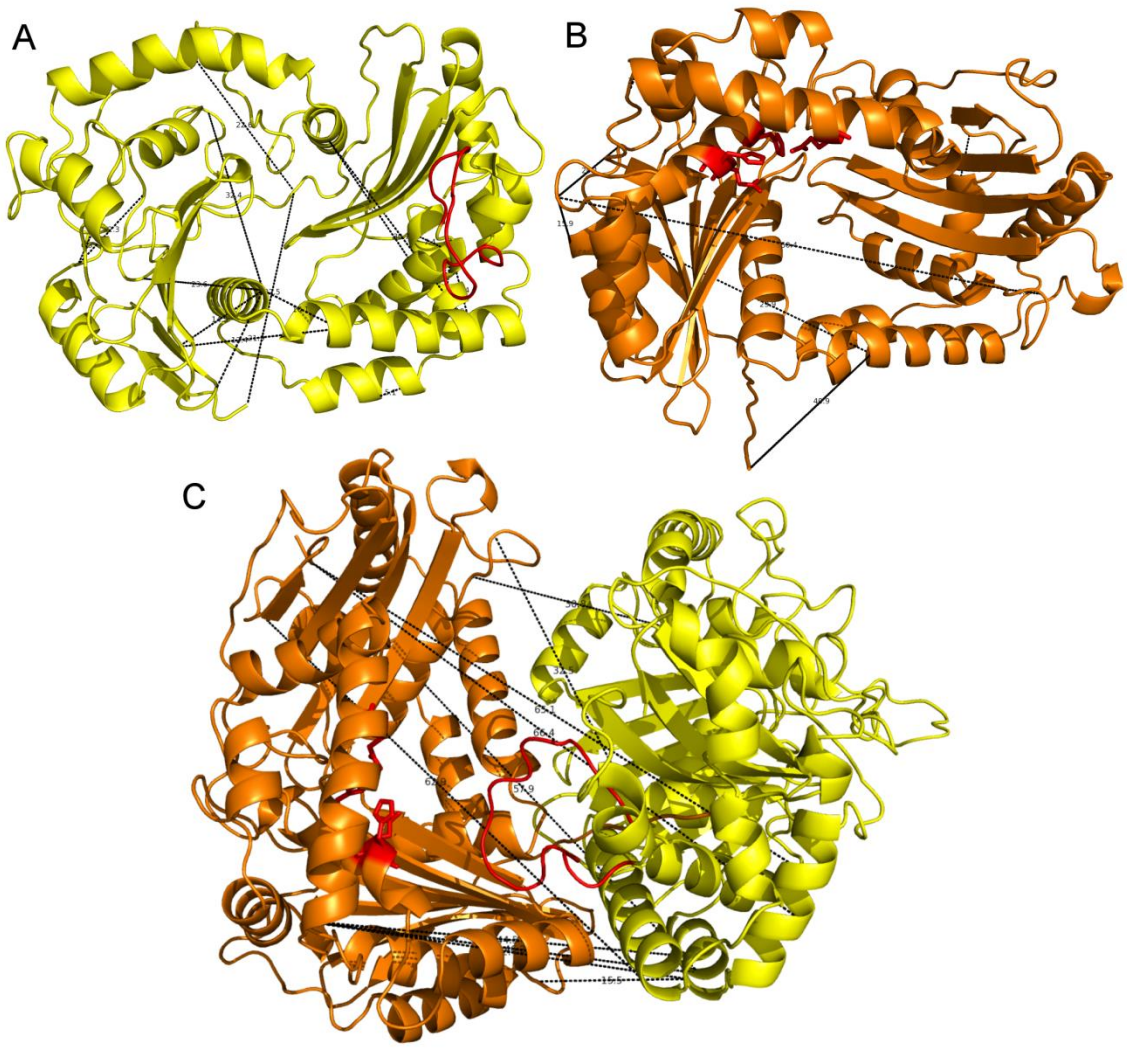


Figure 31. Projection of the identified cross-links between lysine residues on the homology model of HPP. Panel A, B and C show 18, 8 and 10 distances between Cα atoms of lysine residues on α-HPP (yellow), β-HPP (orange) subunit and whole HPP dimer, respectively. Zinc-binding motif of β-HPP subunit and sequence of α-HPP subunit corresponding to the GRL of MPP are displayed in red color.

The second goal of cross-linking experiments was confirmation of the presence of the peptide substrate in the HPP active site. While no cross-links of the substrate with catalytic β-HPP subunit were found, one cross-link between the substrate and the regulatory α-HPP subunit was identified.

5.6 Molecular dynamics simulations of MPP

5.6.1 Substrate translocation from GRL to MPP active site

Targeted molecular dynamics (TMD) simulation was carried out to examine the translocation of the substrate from the GRL (i.e. from the place of the initial recognition of the signal presequence) to the MPP active site (i.e. to the place where the signal presequence cleavage occurs). Both boundary positions are displayed in **Figure 32**, as *GRL-bound* and *AS-bound structures*, respectively. A peptide derived from the malate dehydrogenase (MDH: residues L^{2}SRVAKRAFSST^{13}; the R-2 motif is underlined) signal presequence was chosen as a model substrate.

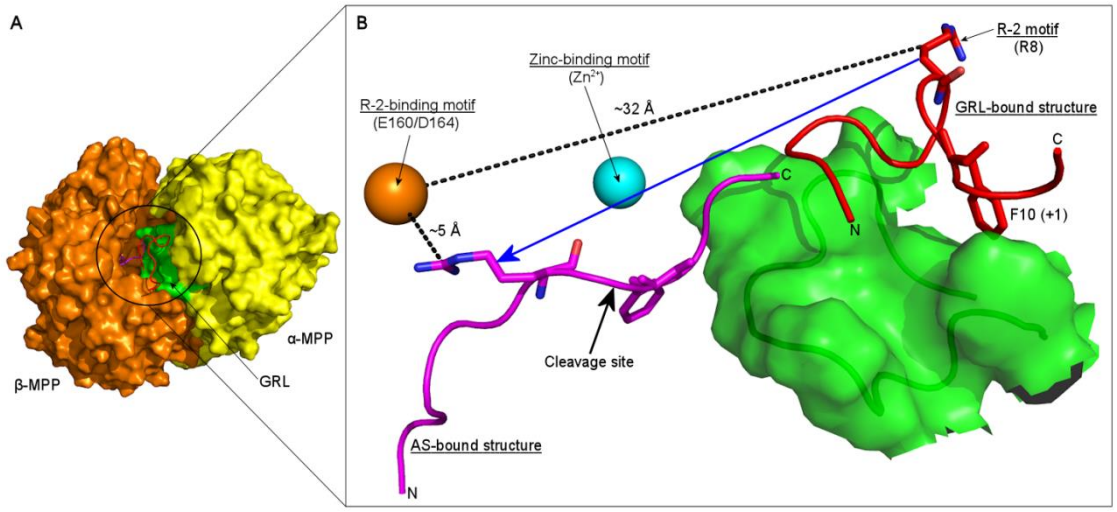


Figure 32. Scheme of substrate translocation from the GRL to the MPP active site. Panel A shows the overall structure of MPP. The van der Waals surface of α-MPP and β-MPP subunits and GRL is in yellow, orange and green, respectively. Panel B shows the conformation and position of the substrate during its recognition by the α-MPP GRL (*GRL-bound structure*; red tube) and just prior to its subsequent proteolysis in MPP active site (*AS-bound structure*; pink tube). The direction of substrate translocation between these two boundary positions is indicated by blue arrow. In the *GRL-bound structure* the substrate residue F10 contributes to the hydrophobic interaction with GRL and the substrate residue R8 (i.e. the R-2 motif) is exposed to the β-MPP subunit. In the *AS-bound structure* the substrate is bound in an extended conformation and its R8 residue interacts with the R-2-binding motif. The zinc-binding motif and R-2-binding motif are shown schematically as cyan and orange spheres which correspond to the zinc-ion and residues E160 and D164 residues of the β-MPP subunit. The distances between the substrate R8 residue and the R-2-binding motif in the *GRL-bound* and *AS-bound structure* are shown as dashed black lines.

We performed three TMD simulations with different restraint durations (**Table 5**) and since all of them provided the same results, for subsequent analysis we selected the longest one, with a restrain period of 1.6 ns and a total simulation period of 1.8 ns (**Video 1** in *Supplementary material*). Two snapshots along the substrate translocation trajectory were chosen for further detailed analysis: one at 0.48 ns, representing a point one-third of the way through the translocation, and a second at 0.84 ns, corresponding to roughly the mid-point of the translocation (*Structure 30-0* and *Structure 50-0* in **Figure 34** and **Figure 35**, respectively). These two structures representing the GRL-substrate interaction at the two selected snapshots were subsequently subjected to a 100-ns-long non-restrained MD simulation to study the GRL-substrate interaction in detail. The MPP dimer remained in its stable, partially-closed conformation during the whole TMD simulation, as documented by an analysis of root-mean-square deviation (RMSD) of Cα atoms (**Figure 33**). The second half of the translocation trajectory was not examined because the substrate was moving through free space, without supporting interactions with the surrounding MPP residues and therefore did not represent a reliable model of the translocation process at this stage.

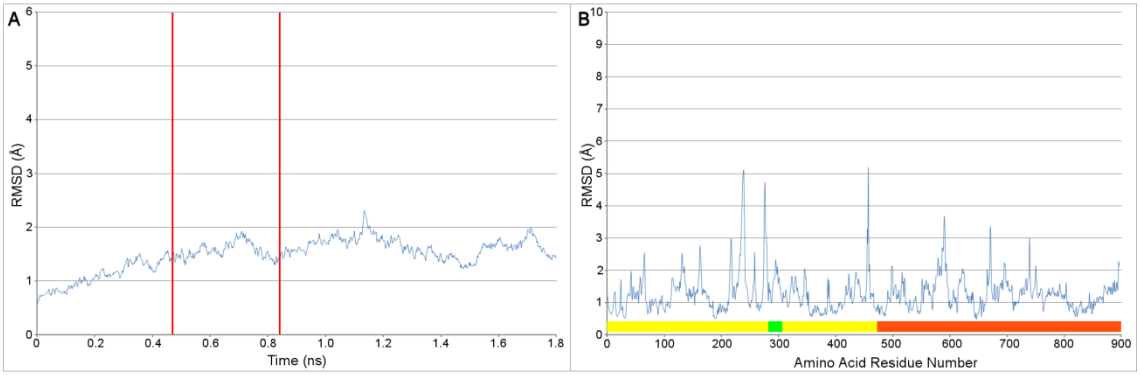


Figure 33. Time-based and residue-based RMSD plots of WT MPP during a TMD simulation of substrate translocation from GRL to MPP active site. A) The RMSD of backbone Cα atoms of WT MPP during a 1.8 ns TMD simulation. The red vertical lines mark the one-third (0.48 ns) and half-way (0.84 ns) points of the trajectory. The structures of these two steps were studied in detail using non-restrained MD simulations. B) The residue-based RMSD of WT MPP at the beginning and the end of the targeted MD simulation. Yellow, orange and green bars along the x-axes indicate the residues corresponding to the α- and β-MPP subunits and to the GRL, respectively. For comparison, in both graphs were used the same scale as those of **Figure 36** and **Figure 38**.

Structure 30-0 in **Figure 34** represents the situation after the substrate has completed the first third of its movement (0.48 ns from a 1.6-ns-long restrain period, i.e. ~30%). Two distances were used to monitor the position of the substrate with respect to the GRL of the α-MPP subunit and the active site of the β-MPP subunit during the subsequent non-restrained MD simulation: (i) the distance between the R-2 motif of the substrate (represented by the ζ-carbon atom of the R8 residue) and the R-2-binding motif of β-MPP (represented by the δ-carbon of the β-subunit E160 residue), and (ii) the distance between the hydrophobic F10 residue in position +1 relative to the substrate cleavage site (represented by its δ-carbon) and the hydrophobic patch of the GRL (represented by the ε-carbon of the α-subunit Y303 residue). These data suggest that the substrate continued to interact with GRL without significant changes in its position and distance from the enzyme active site. In contrast, examining the size of the interaction surface between the substrate and the GRL reveals that the interaction surface decreased from ≈250 Å2 at the beginning of the simulation to ≈180 Å2 at the end (a decrease of 30%). At the beginning of the non-restrained MD simulation, the substrate had no defined secondary structure. Following non-restrained MD simulation, the substrate acquired a stable α-helical conformation and, although it still interacted with the GRL, the N-terminus was reoriented towards the enzyme active site (*Structure 30-100* in **Figure 34**). Although the interaction surface decreased by 30% at this stage the substrate F10 residue still interacted with a hydrophobic patch created by M298 and Y303 of the GRL.

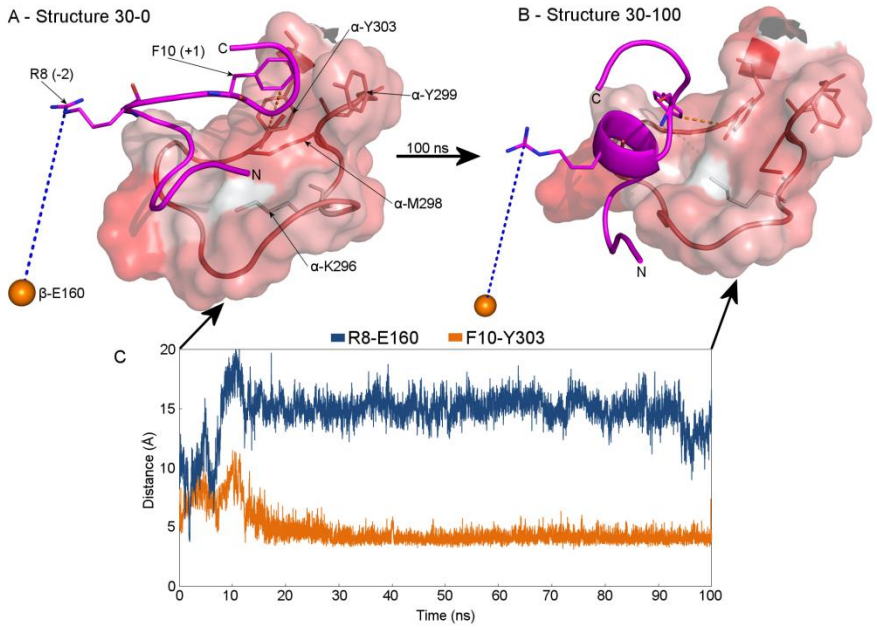


Figure 34. GRL-substrate interaction after the substrate has completed the first third of its translocation. Panel A (*Structure 30-0*) shows a snapshot from the targeted MD simulation corresponding to the situation after 0.48 ns out of a 1.6-ns-long restraint period. Panel B (*Structure 30-100*) shows the GRL-substrate interaction after a 100-ns-long, non-restrained MD simulation performed on *Structure 30-0*. The GRL lies at the entrance to the active site cavity between the α-MPP and β-MPP subunits and is displayed as a semi-transparent van der Waals surface, colored according to residue hydrophobicity [157]. The backbone trace can be seen within this surface and is displayed as a tube. The backbone trace of the substrate is displayed as a magenta tube. Residues K296, M298, Y299 and Y303 of the GRL and residues F10 and R8 (i.e. the R-2 motif) of the substrate are shown as sticks. The α and β prefixes in residue names refer to the α- or β-MPP subunits, respectively. The numbers in brackets show the position of the given residue with respect to the substrate cleavage site. The orange sphere shows the positions of the δ-carbon of the E160 residue and thus represents schematically the R-2-binding motif. The distance between the ζ-carbon atom of the R8 residue and the R-2-binding motif (i.e. δ-carbon of the E160) and the distance between the δ-carbon of residue F10 and the hydrophobic patch of the GRL (represented by the ε-carbon of the α-subunit Y303 residue) are shown as dashed lines in blue ("R8-E160") and orange ("F10-Y303"), respectively. Panel C shows these two distances over the course of the non-restrained MD simulation. Note that (i) the N-terminus of the substrate shifted while the substrate has curled into an α-helix. Moreover, note that during the whole non-restrained MD simulation (ii) the GRL-substrate interaction was stable (the R8-E160 and F10-Y303 distances were largely unchanged) and (iii) the substrate F10 residue interacted with the hydrophobic patch created by GRL residues M298 and Y303.

Structure 50-0 in **Figure 35** represents the situation when the substrate has reached the mid part of its trajectory (0.84 ns of 1.6 ns total, i.e. ~50%). Here, at the beginning of the non-restrained MD simulation the substrate remained in an undefined, partially extended conformation, but still in contact with the GRL

during the first 15 ns of the simulation. Interestingly, later during the simulation, the whole substrate distinctly shifted towards the enzyme active site (*Structure 50-100*). The shift of the substrate is documented by a gradual increase in the F10-Y303 distance during the first half and a subsequent slight decrease during the second part of the simulation period and by a decrease in the interaction surface between the substrate and GRL from ≈120 Å2 at the beginning to 0 Å2 at the end of the non-restrained MD simulation. The substrate R8 residue (i.e. the R-2 motif) was reoriented towards the β-subunit E160 residue, the part of the R-2-binding motif that participates in substrate binding in the MPP active site.

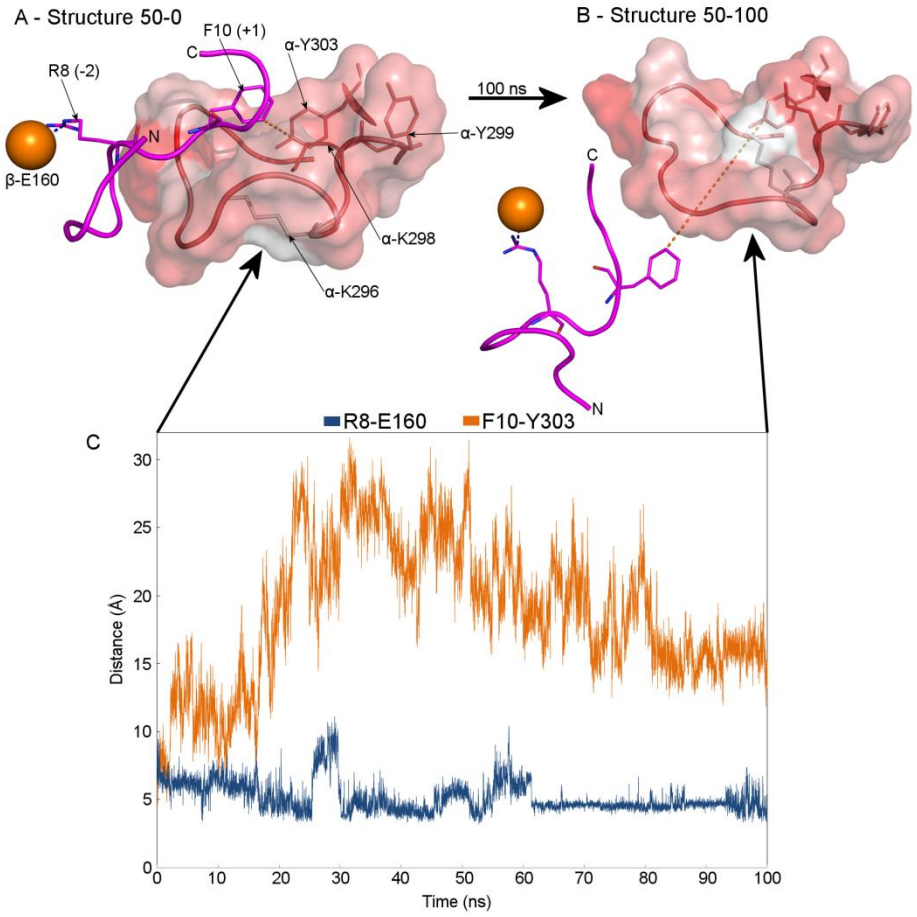


Figure 35. GRL-substrate interaction after half the substrate translocation. Panel A (*Structure 50-0*) shows a snapshot from the targeted MD simulation corresponding to the situation at time 0.84 ns of a 1.6-ns-long restraint period. Panel B (*Structure 50-100*) shows the GRL-substrate interaction after a 100-ns-long non-restrained MD simulation performed on *Structure 50-0*. The structural elements are represented as in Figure 2. Note that (i) the substrate has now shifted completely away from the GRL towards the MPP active site, (ii) the R8-E160 distance decreased from 8 to 4 Å and that (iii) the R8 residue has reoriented towards the R-2-binding motif.

5.6.2 Structural role of GRL for the MPP dimer stability

An all-atomic, non-restrained MD simulation with explicit solvent was carried out to gain insights into the structural dynamics of MPP. The structure of yeast wild-type MPP (WT MPP) with bound substrate in its active site was taken from the Protein Data Bank. The substrate is derived from the CytC oxidase IV (COX IV: residues S^{7}IRFFKPAT<u>R</u>T^{17}; the R-2 motif is underlined) signal presequence. The unbound structure of WT MPP was produced by removing the substrate from peptidase active site. Analogously, two models of mutant MPP with deletion of the GRL (ΔGRL MPP) were constructed - a ΔGRL MPP with the substrate bound in its active site and a ΔGRL MPP with no bound substrate.

The analyses of RMSDs of MPP backbone Cα atoms were chosen as a tool for monitoring the trajectory stability and conformational changes (**Figure 36**). The 2D representations of RMSD are shown in **Figure 37**. The systems representing WT MPP with and without a bound substrate showed stable conformations during the entire production phase of the simulations (**Figure 36-A**). After 10 ns, the RMSD of WT MPP fluctuated slightly around a mean value of 2.5 Å (±0.5 Å). Binding of the substrate to the MPP active site seemed to stabilize the MPP conformation, however, since the overall backbone RMSD decreased by 0.7 Å during the first 30 ns of simulation and gradually rose to a mean value of 2.5 Å at the end of the simulation. Examining the size of the interaction surface between the two MPP subunits reveals that it increased from ≈2100 Å2 at the beginning of the simulation to ≈2300 Å2 at the end (**Table 9**). In contrast, the unbound structure showed the opposite trend, with the interaction surface decreasing from ≈2100 Å2 at the beginning to ≈1800 Å2 at the end. (It is important to note that these numbers include only the contact area between the two subunits and have no contribution by the substrate.) These data suggest that binding of the substrate in the WT MPP active site strengthens the dimer interface.

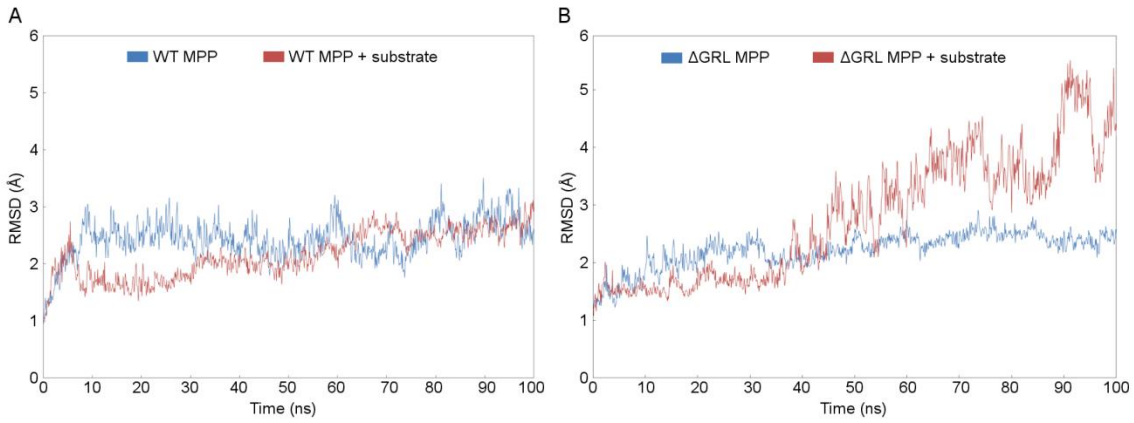


Figure 36. Time-based RMSD of backbone Cα atoms of WT MPP and ΔGRL MPP. Panel A shows the change in the RMSD of the WT MPP structure over the course of a 100 ns simulation with respect to the initial model. Runs both with (red) and without (blue) a active site-bound peptide substrate are shown. Panel B shows the same information, but for the ΔGRL MPP structure. The ΔGRL MPP structure was produced by deleting residues 285-300 from the α-MPP structure. **Figure 37** displays the same information in 2D RMSD plots.

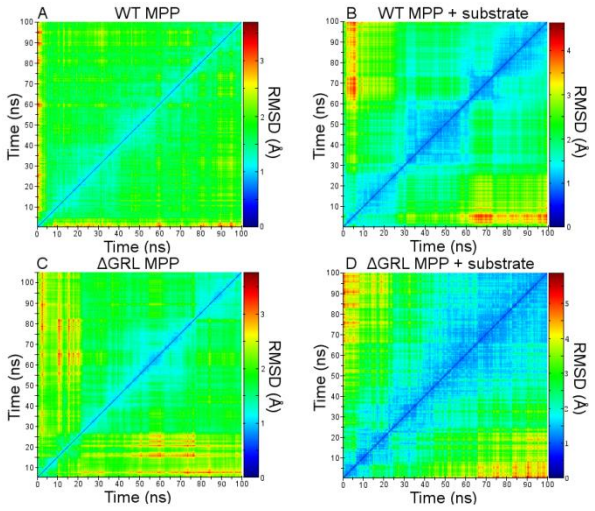


Figure 37. 2D plots of backbone Cα RMSDs during non-restrained MD simulations of WT and ΔGRL MPP. Panels A and B show 2D plots of the change in the RMSD of the Cα atoms of the WT MPP structure over the course of a 100 ns simulation with respect to the initial model. Panel A shows WT MPP without a bound substrate and panel B shows the active site-bound form. Panels C and D show the same information for the free (C) and bound (D) forms of ΔGRL MPP.

	α-MPP versus β-MPP		Substrate versus β-MPP	
	Start value	**End value**	**Start value**	**End value**
WT	2110 Å²	1812 Å²	-	-
WT + substrate	2110 Å²	2336 Å²	886 Å²	886 Å²
ΔGRL MPP	1850 Å²	1907 Å²	-	-
ΔGRL MPP + substrate	1850 Å²	1428 Å²	886 Å²	861 Å²

Table 9. Interaction surfaces between the α and β-subunits of MPP and between the β-subunit and the bound peptide substrate in MPP active site.

In the absence of the substrate, the ΔGRL MPP enzyme structure did not undergo significant conformational changes. This is illustrated by the fact that its backbone RMSDs fluctuated around a mean value of 2.5 Å for the second half of the MD simulation (**Figure 36-B**). These fluctuations are smaller than for the WT MPP since the loss of large flexible GRL reduced the overall variability. On the contrary, binding of the substrate appeared to destabilize the ΔGRL MPP dimer. During the MD simulation, the RMSD didn't reach a plateau and instead was continuously increasing, with a RMSD value of 4.5 Å at the end of simulation, almost double that observed in the WT MPP simulation. There were also wide fluctuations of about 1 Å. Examining the amount of surface area per monomer buried on dimerization shows that it increased slightly from ≈1850 Å2 to ≈1900 Å2 for the model lacking bound substrate and decreased substantially from ≈1850 Å2 to ≈1400 Å2 for the model with bound substrate (**Table 9**).

During both WT MPP and ΔGRL MPP simulations, the substrate remained bound in the MPP active site, as shown by an assessment of the interaction surface area between the substrate and the β-MPP subunit (**Table 9**). In addition, RMSD of Cα atoms of the substrate was calculated between the structures at the beginning and the end of MD simulations. RMSD of the full-length substrate was 2.5 Å and, naturally, the most flexible parts were the C- and N-terminus of the substrate. When one or two residues were removed from each termini of the substrate, the RMSD decreased and reached 1.2 Å and 0.5 Å, respectively.

To more directly compare WT MPP and ΔGRL MPP, a residue-based RMSD analysis was used which indicated that the differences in backbone flexibility arise from the presence of the substrate in the enzyme site (**Figure 38**). For WT MPP, the presence of the substrate did not affect the enzyme structure significantly, though a small decrease in the β-MPP RMSD might have been present (**Figure 38-A**). On the other hand, for ΔGRL MPP, the presence of substrate in the enzyme cavity caused a large increase in the RMSD, indicating that the enzyme structure was significantly affected (**Figure 38-B**). In general, the most flexible parts of both MPP structures are their surface loops and N- and C-termini. The deletion of the GRL destabilized those parts of the MPP subunits that are not in direct contact and farthest from the dimer interface (**Figure 38-C** and **Figure 38-D**). Thus, the ΔGRL MPP dimer with bound substrate appears to be more open and less stable than the WT structure.

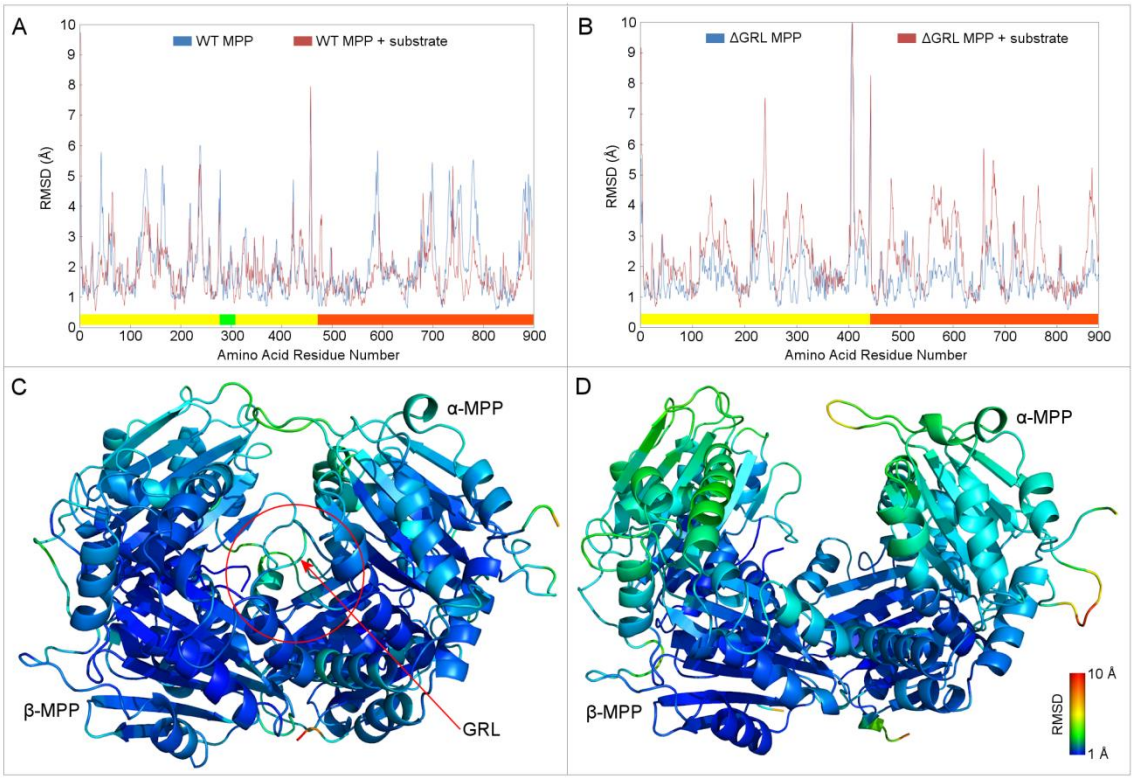


Figure 38. Residue-based RMSD of backbone Cα atoms of WT MPP and ΔGRL MPP. Panels A and B show the RMSD per residue of the WT MPP and ΔGRL MPP structures, respectively. The red lines show the RMSD of the structures in the presence of the active site-bound substrate and the blue lines represent the unbound structures at the end of a 100-ns-long MD simulation. The bars

along the x-axes indicate the residues which belong to the α (yellow) and β-MPP (orange) subunits, while the green bar in panel A indicates the position of the GRL (residues 285-300) which was deleted in the ΔGRL MPP structure. Panels C and D show, respectively, the WT MPP and ΔGRL MPP structures with active-site bound substrate colored to reflect their per-residue RMSDs and structurally aligned according to the parts with the lowest fluctuation of RMSD. Both structures are colored according to RMSD scale bar in the bottom right corner of panel D. The red circle marks the position of the GRL. Note that the ΔGRL MPP dimer appears to be more open than the WT structure and that the RMSDs of the areas farthest from the dimer interface are notably higher.

A radius of gyration analysis supports this conclusion (**Figure 39**). According to this analysis, when the substrate binds to the active site of WT MPP, the radius shrank, indicating that the enzyme dimer had become more compact. In contrast, the structure of ΔGRL MPP bound to a peptide substrate had a larger radius of gyration which increased during the course of the simulation. This indicates that the dimer became looser or more open and less stable.

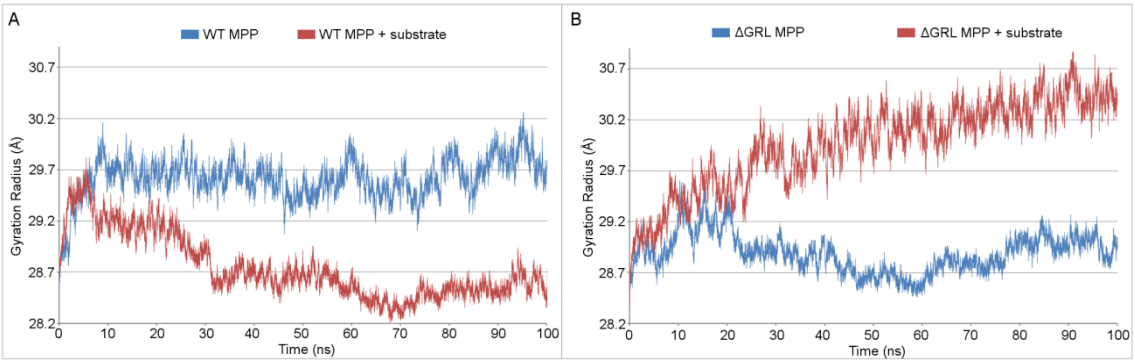


Figure 39. Radius of gyration of WT MPP and ΔGRL MPP without and with bound substrate in the MPP active site. Panel A shows the radius of gyration of the WT MPP structure over the course of the 100 ns simulation both with (red line) and without (blue line) a bound peptide substrate. Panel B shows the same information for the ΔGRL MPP structure. Note that for the WT structure, the substrate causes the radius to shrink while it increases greatly for the ΔGRL MPP structure.

6. DISCUSSION

6.1 Mitochondrial processing peptidase

Our previous results, obtained using tryptophan fluorescence measurements and MD simulations, showed that the GRL of α-MPP is the site where the primary interaction between the signal presequence (as a substrate) and MPP occurs, and suggested that the substrate's α-helical conformation was important for this interaction [118]. In this study we employed MD simulations to study the role of the GRL in the process of substrate translocation from the GRL to the MPP active site and also its role in the tertiary and quaternary structures of MPP.

6.1.1 GRL as an active element during substrate translocation

The large number of glycine residues, together with the weak electron density for the GRL in the MPP crystal structure, indicates that this loop is highly flexible. Using a targeted MD simulation, we simulated the process of substrate translocation from the GRL to the MPP active site (**Figure 32** in *Results*). During this process, the GRL undergoes significant conformational changes in the part containing residues 289-293 in α-MPP (**Video 1** in *Supplementary material*). However, these changes did not appear to affect mutual MPP subunit interactions, since RMSD of the backbone Cα atoms fluctuated around an average value of 1.5 Å (**Figure 33**) and interaction surface between the two MPP subunits fluctuated between values 2150 and 2250 Å^2 (**Table 10**) during the whole simulation. We chose two steps along the trajectory and studied the enzyme-substrate interaction by non-restrained MD simulation.

Time	α-MPP versus β-MPP
0 ns	2250 Å^2
0.3 ns	2290 Å^2
0.6 ns	2150 Å^2
0.9 ns	2200 Å^2
1.2 ns	2110 Å^2
1.5 ns	2120 Å^2
1.8 ns	2200 Å^2

Table 10. Interaction surfaces between the α- and β-subunits of MPP during TMD simulation.

First we studied the enzyme-substrate interaction when the substrate had finished the first third of its trajectory (**Figure 34**). To our surprise, the substrate, which at this stage had no definable secondary structure, acquired at the end of the following non-restrained MD simulation an α-helical loop stabilized by a hydrophobic interaction between hydrophobic residues on one side with a hydrophobic patch of the GRL and by hydrogen bonds between positively charged arginine residues on the other side of the α-helix with negatively charged residues on the β-MPP subunit. Furthermore, the process of α-helix loop folding was accompanied by reorientation of the N-terminus of the substrate towards the enzyme active site. Thus, although initially we hypothesized that hydrophobic interactions play an important role only during the initial substrate recognition [118], now it seems to be likely that hydrophobic interactions take significant part also later during the substrate translocation process.

Subsequently, the substrate reached the mid part of its translocation trajectory (**Figure 35**). At this stage the substrate R8 residue (i.e. the R-2 motif) did not interact with either the E160 or D164 residues of β-MPP (i.e. the R-2-binding motif), as they did in the MPP crystal structure. However, during the following non-restrained MD simulation the substrate shifted notably deeper towards the MPP active site and the R-2 residue was reoriented in several steps towards the E160 residue of the R-2-binding motif. The position of R8 at different stages of substrate translocation is shown schematically in **Figure 40**.

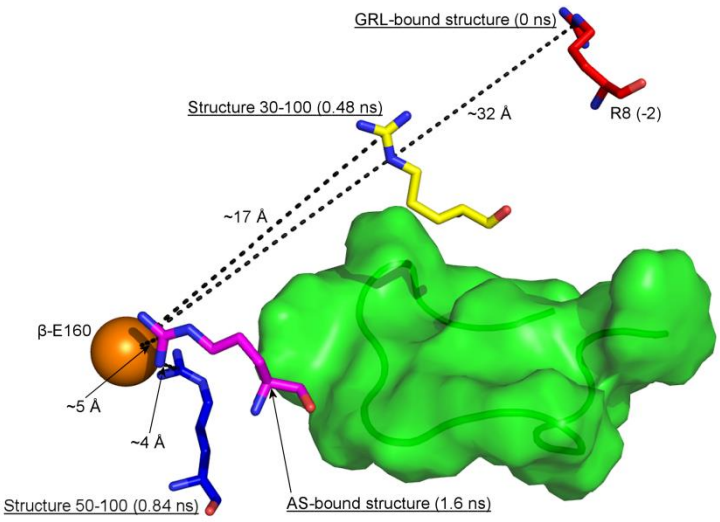


Figure 40. Scheme showing the positions of the R8 residue along the substrate translocation trajectory. The R8 residue is located in position -2 relative to the substrate cleavage site and thus represents the R-2 motif. The R-2-binding motif is shown schematically as an orange sphere representing the δ-carbon atom of the β-subunit E160 residue and the GRL is displayed as a green semi-transparent van der Waals surface whose backbone chain is displayed as a tube. In the *GRL-bound structure* the substrate is bound to GRL and residue R8 (red sticks) is exposed to the β-MPP subunit. In the *AS-bound structure* (purple sticks), on the other hand, the substrate is bound in the MPP active site and the R8 residue interacts with the R-2-binding motif. The positions of the R8 residues in *Structure 30-100* (yellow sticks) and *Structure 50-100* (blue stick) were obtained at the end of a 100-ns-long non-restrained MD simulation performed on the structures corresponding to the snapshots from a targeted MD simulation when the substrate reached the first third and the mid part of its translocation trajectory. The numbers in brackets mark the time steps along the substrate translocation trajectory. The distances between the R8 residue and the R-2-binding motif are shown as dashed black lines. Note that while *Structure 30-100* has the R8 residue oriented away from the GRL, in *Structure 50-100* it is oriented towards the R-2-binding motif and its distance from it is almost the same as in the *AS-bound structure*.

The tendency of the substrate to shift spontaneously towards the MPP active site is illustrated by the lengthening of the distance between the substrate F10 and GRL Y303 residues and the simultaneous approach of the substrate R8 residue (i.e. the R-2 motif) to the R-2-binding motif of the β-MPP subunit (**Figure 35-C**). An interaction surface analysis between the substrate and the GRL further confirms this tendency, which becomes even more evident later after the substrate has reached the mid part of its translocation.

The role of the GRL has also been studied by site-directed mutagenesis. Nagao et al. reported that MPP with mutations F289A, F289L, K296A or M298A in

the GRL of the α-MPP subunit had 10-fold less affinity for substrate peptides than did the wild-type and that their activities decreased to 1% [16] [102]. However, the mutation M298L decreased the activity by only half, suggesting that the important feature of the amino acid residue in this position is its hydrophobic character. Mutations of the partially hydrophobic residue Y299 to serine and alanine had the same effect. The authors used a pull-down assay to investigate the effect of these mutations on the subunit-subunit interaction and suggested that the decrease in activity is not due to an incorrect or insufficient interaction between the MPP subunits. Our findings suggest, however, that the subunit interactions may be weakened as a result of these point mutations, though perhaps they remain strong enough to allow the MPP subunits to interact during the pull-down assay. Moreover, we conclude that K296 of the GRL appears to act as a stabilizing element for the whole GRL: In the absence of a substrate, it restricts the flexibility of the GRL and therefore controls the size and shape of the hydrophobic patch created by its side chain together with the M298, Y303 and Y299 residues.

6.1.2 GRL keeps MPP dimer in a partially closed conformation

A highly conserved MPP has been found in the mitochondria of different species, including yeast, mammals, plants and protozoa. Moreover, homologs of MPP have recently been found in hydrogenosomes [130], the organelles evolutionary-related to mitochondria. Specifically, a hydrogenosomal processing peptidase (HPP), a MPP-like protein, was characterized from the hydrogenosomes of the human parasite *Trichomonas vaginalis* [131]. MPP-like proteins are also present in bacteria [141] and the ancestral bacterial peptidase similar to the α-proteobacterial *Rickettsia prowazekii* peptidase (RPP) is considered as a progenitor of the modern MPP [129]. Several other bacterial peptidases have recently been described [100, 103], including an example of the heterodimeric M16 peptidase from α-proteobacterium *Sphingomonas sp.* [139]. This one will be discussed in more detail below.

A common trait of all these peptidases is the presence of a partial or complete GRL-like structural element, which may influence their quaternary structure and therefore their mechanism of action. Although the GRL-like element of HPP differs from the GRL of α-MPP in the sequence (**Figure 12**), the GRLs of all eukaryotic MPPs are highly conserved, indicating that they are vitally involved in the protein's biological function. The bacterial peptidases, on the other hand, contain just "embryonal" GRL, if any. An example is shown in **Figure 41**.

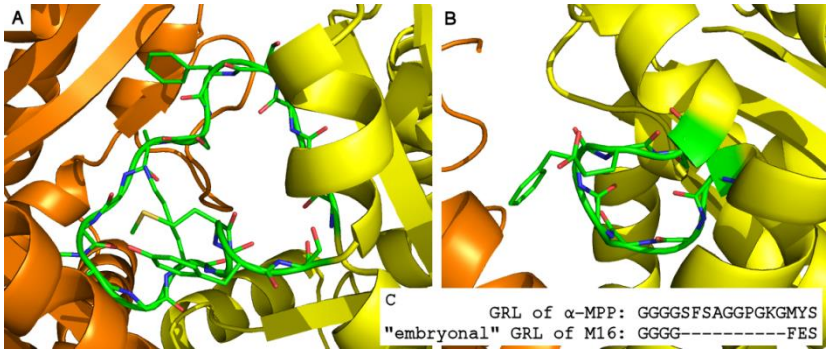


Figure 41. Detail structures of GRLs of MPP and M16 peptidase from *Sphingomonas sp.* Panel A shows the GRL region of MPP while panel B shows the "embryonal" GRL of M16 peptidase. The α-subunit is shown in yellow, β-subunit in orange, and the GRL in green. The side chains of amino acids 285–300 of the MPP GRL and 290–296 of the M16 peptidase GRL are represented as sticks. Panel C shows a sequence alignment of both GRLs.

The effect of the GRL on MPP tertiary and quaternary structure was studied by non-restrained MD simulations performed on WT and ΔGRL MPP models, both with and without a peptide substrate bound in the active site. Analyses of RMSD of the backbone Cα atoms of MPP structures with active-site bound substrate, the interaction surface and the radius of gyration showed that while the WT MPP became more compact, ΔGRL MPP acquired a less compact and more open conformation.

Several experimental studies have been published which focus on the GRL. A pull-down assay of various α-MPP GRL deletion mutants by β-MPP-His6 revealed that these subunits continued to associate stoichiometrically, leading the authors to suggest that the GRL does not influence the association between subunits [102]. Furthermore, an MPP dimer containing an α-MPP with residues

249-287 deleted was still able to cleave the malate dehydrogenase (MDH) presequence, albeit with lower processing activity than that of WT MPP, but could process longer presequences only inefficiently [129]. On the other hand, the deletion of G292 makes α-MPP alone unable to bind (i.e. recognize) the short presequence of yeast MDH [118].

We are proposing a unifying interpretation of these contradictory experimental results. We suggest that GRL is a crucial structural element, which is responsible for holding the MPP binding cleft in a finely adjusted, partially closed conformation. Larger deletion of this part does not prevent the MPP subunits from association, but in the presence of substrate it does cause this cleft to adopt a more open dimer conformation, which may alter the peptidase specificity, but preserve the partial peptidase activity. On the other hand, the deletion of only one amino acid may change the conformation of GRL itself and thus "close the entrance" to the MPP active site. The normal, partially closed conformation of WT MPP is not affected by the presence or absence of a substrate in its active site, unlike *Sphingomonas sp.* M16 peptidase, the peptidase that lacks a GRL, which adopts distinct closed and open conformations, depending on whether there is or is not a substrate bound to its active site [139].

6.2 Hydrogenosomal processing peptidase

The first biochemical characterization of HPP was carried out by Brown et al. who suggested that HPP is functional as homodimer of β-HPP subunits [130]. On contrary, later we proved that HPP is functional as heterodimer, consisting of α- and β-HPP subunit, similarly to MPP. The first indication for such an organization was the presence of a gene coding for α-HPP subunit in *T. vaginalis* genome. Although the overall homology of α-HPP with α-MPP subunit is very low (~17% of identity and ~26% similarity), one short sequence of length of ~20 residues showed remarkably high similarity with GRL of α-MPP subunit. The subunit interaction was proved by pull down assay and by proteolytic activity assay using substrate peptide labeled with fluoresce dye. Another indication was yeast two-

hybrid system that not only confirmed the α-β interaction but also suggested that β-HPP subunits are capable of forming homodimers, although very likely only transient and not stable [143]. However, more detailed information on HPP structure was missing.

The knowledge of the HPP structure would help to answer numerous questions, such as how the differences in structure of HPP and MPP, especially in the part of GRL, determine the specificity of corresponding peptidases? Or what is the evolutionary relation between these two peptidases and the other M16 family members? Hence, apart from homology model construction, the decision for an attempt to solve the three-dimensional structure of HPP was made. To accomplish this goal, it was crucial to optimize the purification procedure of HPP.

6.2.1 Purification procedures

A number of purification strategies were carried out before figuring out the optimal purification process. The first approach was the separate expression of WT α-HPP and β-HPP with C-terminally attached His-tag [143]. This approach was suitable for initial activity assays however the sample quality was not sufficient for crystallization experiments since the subunit composition was not equimolar – β-HPP subunit was always in excess. Later, the system for co-expression of both subunits (two pET expression vectors in one cell) was worked out but resulted in the same quality of HPP dimer. Further, α- and β-subunits were cloned into pET Duet vector – the expression in this system did not work at all since only β-subunit was produced.

After unsuccessful purification experiments of HPP dimer with C-terminally His-tagged β-HPP and N-terminally His-tagged α-HPP subunit, expression vector for α-HPP subunit with C-terminally attached His-tag was prepared. Although the majority of α-subunit in this form was expressed also in inclusion bodies, a reasonable amount was expressed in the native form suitable for subsequent metal affinity purification together with WT β-HPP subunit.

GroES/EL shaperonins showed to have a positive effect on the yield of α-subunit produced in native form and in inclusion bodies. On contrary, different host cell strains (BL21(DE3) and Rosetta *E. coli*) did have an effect on this yield.

In the next step, purification procedure for HPP was optimized. In this part it is worth mentioning the importance of glycerol in the purification procedure since no other buffer compound showed so great importance such as glycerol. Glycerol is known to shield protein-protein hydrophobic interaction through binding to the protein hydrophobic patches. Indeed, when a buffer containing only 5 or 10% of glycerol was used during washing of the metal affinity column with bounded HPP dimer, the subunit composition of the final HPP sample was not equimolar (again, β-subunit was in excess). On contrary, when a buffer with 20% of glycerol was used proper HPP dimer was obtained. This finding suggests that β-subunit is capable of a nonspecific interaction with another β-subunit which is in good agreement with our previous results obtained using yeast two-hybrid system and with other experiments, as will be discussed later.

In addition to the SDS-PAGE, DLS method was employed to check the sample homogeneity. We have shown that HPP dimer purified as discussed above and stored at -20 °C or -80 C° is stable for at least one month and has the same activity as freshly prepared sample. Although HPP is stable both at -80 °C and -20 °C, the samples were routinely stored at -80 °C. The DDM detergent was shown to have a great positive effect on HPP dimer stability and was used routinely.

6.2.2 Structural features of HPP

The first step towards understanding the HPP structure was construction of homology model based on the available crystal structure of MPP. The model was built despite the fact that the homology between HPP and MPP subunits is below the limit recommended for homology models construction (i.e. 30%) and, thus, could be interpreted only with caution. The model was later validated

experimentally, as will be discussed later. However, the ultimate information about HPP organization would provide three-dimensional crystal structure.

Crystallization experiments were performed both on fresh as well as on frozen samples of WT form of HPP in the absence of peptide substrate. Although 6 different crystallization screens in total were tested, we didn't succeed in growing proper protein crystal. Later the E56Q proteolytically inactive mutant form of HPP was prepared for intended crystallization experiments in the presence of peptide substrate. Such a type of experiments is currently being performed.

HPP is homologous not only with MPP but also with other peptidases of the M16B family. While MPP subunits resemble partially open-closed conformation regardless the presence of the substrate in its active site, some M16B peptidases adopt two distinct conformations depending on the presence of active-site bound substrate. Therefore, we were interested in whether the HPP conformation is affected by substrate or nor. In order to get insights into the HPP structure a few advanced techniques were employed along with the crystallization experiments. Specifically, three methods were used - biological small-angle X-ray scattering, hydrogen-deuterium exchange and chemical cross-linking studies.

All three methods suggest that HPP does not undergo conformation changes in the presence of substrate, or the changes are minor and are not detectable by those methods. However, the question is if the substrate is really present in the peptidase active site during experiments. When answering this question, following points must be kept in mind: (i) peptide substrate is derived from adenylate-kinase signal presequence verified experimentally [143], (ii) HPP cleaves the substrate (MALDI-TOF substrate cleavage assay) and (iii) one cross-link between the substrate and the regulatory α-HPP subunit was identified. On the other hand, (i) no cross-link between the substrate and the catalytic β-HPP subunit was identified and (ii) H/D exchange did not show changes in deuteration kinetics upon substrate addition to the sample. The fact that substrate-β-subunit cross-links were not identified may be simply explained on the basis

that in the vicinity of bound substrate there are no lysine residues present in suitable distance for cross-link formation. With respect to the H/D experiments it is possible that the substrate binding and releasing is very fast and could not be registered as a change in deuteration kinetics curves. In the future it would be interesting to repeat these experiments with a specific peptidase active-site-bound inhibitor. To sum it up with regard to possible conformation changes specific for the moment of peptidase activity, HPP keeps one conformation and, thus, behaves similarly to homologous MPP. In addition, bio-SAXS experiments performed on MPP confirmed that MPP behaves in this respect in the same manner. Therefore, it is very likely that HPP subunits resemble partially open-close conformation, similarly to the homologous MPP.

Surprisingly, bio-SAXS experiments demonstrated a general conformation change between WT and E56Q HPP and subsequent H/D exchange experiments refined the original findings. Although the single E56Q mutation in β-HPP subunit affects the conformation of whole HPP dimer the major changes were registered on β-HPP subunit. Interestingly, WT form of β-HPP exhibits higher H/D exchange rates that E56Q mutant form, suggesting that mutant β-subunit is more stable than WT form and resists more the H/D exchange. In this respect our findings are in good agreement with MPP crystal structure that was solved both for WT and E73Q mutant form of MPP [88] where the authors noted that the mutant and WT structures of MPP without peptide substrate have a RMSD of 0.14 Å for all Cα atoms and 0.37 Å for all atoms of zinc-binding motif (βHis-70, βGlu-73, βHis-74 and βGlu-150)[10]. The absence of the zinc-ion in the HPP structure has the same effect as E56Q mutation, however not so significant.

Further we investigated the solution-shape of HPP using bio-SAXS experimental data and *ab-initio* modelling techniques. Solution-shapes were

[10] The E56Q mutation causes proteolytic inactivity of HPP and was chosen based on the E73Q mutation in MPP since both glutamate residues correspond to the same residue in the conserved zinc-binding motif.

calculated both for WT and E56Q mutant form of HPP. In both cases, a slightly elongated shape was found to fit well the experimental data. In the next step, the homology model was docked into solution-shape structures and in all cases the homology model fitted well the globular part of the solution shape structures. However, an extra "tail extensions" were observed and we can just speculate if the "tail extension" is a native part of the HPP structure or if it is just artifact arisen during data collection and/or processing. The fact that similar extensions were found also in the case of MPP solution shapes would suggest the later. The definite answer may provide only three-dimensional crystal structure of HPP.

Cross-linking studies were employed to validate the homology model of HPP. As was mentioned earlier, the model was built on the crystal structure of MPP while the homology of both enzymes is below the limit generally recommended for homology models construction. In this respect the results of cross-linking experiments are not surprising - the validity of HPP model is very limited. Specifically, although the validity of homology models of individual HPP subunits (i.e. tertiary structure) is low but acceptable, the quaternary structure of homology model of HPP dimer as a whole is not reliable at all. Interestingly, the cross-linking studies suggested that β-subunit might be capable of forming "head-to-tail" homodimers which is in good agreement with purification experiments with His-tagged α-HPP subunit where it was impossible to obtain equimolar HPP dimer when a washing buffer with less than 20% of glycerol was used (β-subunit was always in excess).

Further, the second goal of cross-linking studies was confirmation of the presence of the peptide substrate in the active site of proteolytically inactive E56Q form of HPP. The peptide substrate was derived from adenylate-kinase presequence and contained three lysine residues (MLSTLA**K**RFASG**KK**DRM). The expected cross-links between the substrate and catalytic β-HPP subunit were not found. Surprisingly, such a cross-link was found between K14 substrate residue and K197 reside of regulatory α-HPP subunit. On the other hand, this finding

doesn't need to be necessarily surprising considering the fact that the active site of homologous MPP is formed by both MPP subunits while both subunits also participate in the process of substrate recognition and binding. The distance between Cα atoms of those cross-linked lysine residues was not possible to measure since the substrate is not present in homology model of HPP. The fact is that the K197 residue is present on the surface of homology model and the Cα distance would be >25 Å and would not fulfill the cutoff criterion. Thus, in this respect the results are ambiguous.

6.3 Evolution of MPP - dawn and fall of GRL

The goal of this part of discussion is to introduce an evolutionary scenario for GRL of MPP and its physiological function in the context of the eukaryotic cell evolution.

The presence of mitochondrion is one of the ultimate features distinguishing eukaryotic cells from bacterial ones. Although the origin of mitochondrion was successfully traced to α-proteobacteria, the process of mitochondrion acquisition is still matter of discussion. One of the most critical steps in creating mitochondria (i.e. the transition from autonomous endosymbiont to organelle) was the protein import machinery establishment in the membrane of what originated from bacterial symbiont and the related process of genome reduction. From this point of view, the two principal hypotheses exist – "insiders'" and "outsiders'" model.

The validity of the "outsiders'" model denotes the presence of Tom20 receptor which does not have bacterial homolog or dual function of protein translocation machines Tim23/Tim17 [47]. However, some components of protein translocation machinery are clearly of bacterial origin and thus validate also the "insiders'" model. Illustrative examples of components of bacterial origin may be highly conserved inner membrane protease (IMP), Tim44 or Tom50.

The question is which evolutionary scenario was following MPP? Genuinely, MPP does not have bacterial counterparts since its processing activity towards signal presequences in bacteria is needless. Therefore, it might seem that MPP evolution should fall in the category of "outsiders'" model and that MPP was established *de novo* after the establishment of mitochondria and associating protein translocation mechanisms across mitochondrial membranes. However, we are going to present a hypothesis showing that the "insiders'" model might be more suitable for our understanding to the evolutionary origin of MPP.

MPP belongs to the M16B family of peptidases which consists of eukaryotic as well as of bacterial peptidases. BLAST search combined with search through the PDB database of protein structures using MPP structure as a probe revealed a couple of interesting bacterial peptidases. The last three in the list were found in PDB database without references to publications:

- **RPP –** *Rickettsia prowazekii* peptidase. Monomeric zinc-binding motif containing peptidase with cleavage preference for basic peptides. Crystal structure not known. [129]

- **BHP –** *Bacillus halodurans* peptidase. Homodimeric peptidasome with known crystal structure – every subunit contains zinc-ion binding and RY motifs (**Figure 42** - A). In addition, two adjacent homologous genes were found in the genome very likely coding for heterodimeric peptidase – one subunit contains zinc-binding motif (BH2392) while second RY motif (BH2393) [100, 138].

- **TTHA1264 –** *Thermus thermophilus* peptidase. Crystal structure of homodimeric form (**Figure 42** - B) is very likely artificial since adjacent pair of homologous genes was later found in the genome. Thus, native TTHA forms very likely also heterodimeric peptidase – one subunit contains zinc-binding motif while second RY motif. [100, 103]

- **SPH –** *Sphingomonas sp.* peptidase. Crystal structure of heterodimeric peptidase in two conformations – open and close (**Figure 42** - C). One subunit contains zinc-binding motif while second RY motif. [139]

- **PDB ID: 3GWB**. Peptidase from γ-proteobacterium *Pseudomonas fluorescence*. The structure contains RY motif and "embryonal" GRL but no zinc-binding motif (**Figure 42** - D).

- **PDB ID: 3D3Y**. Peptidase from eubacterium *Enterococcus faecalis*. The structure contains RY motif and "embryonal" GRL but no zinc-binding motif (**Figure 42** - E).

- **PDB ID: 3IVL**. Peptidase from β-proteobacterium *Bordetella parapertussis*. Although the structure is not complete it has structural signs of the β-subunit and the sequence contains zinc-binding motif (**Figure 42** - F). In the structure is visible "embryonal" form of GRL.

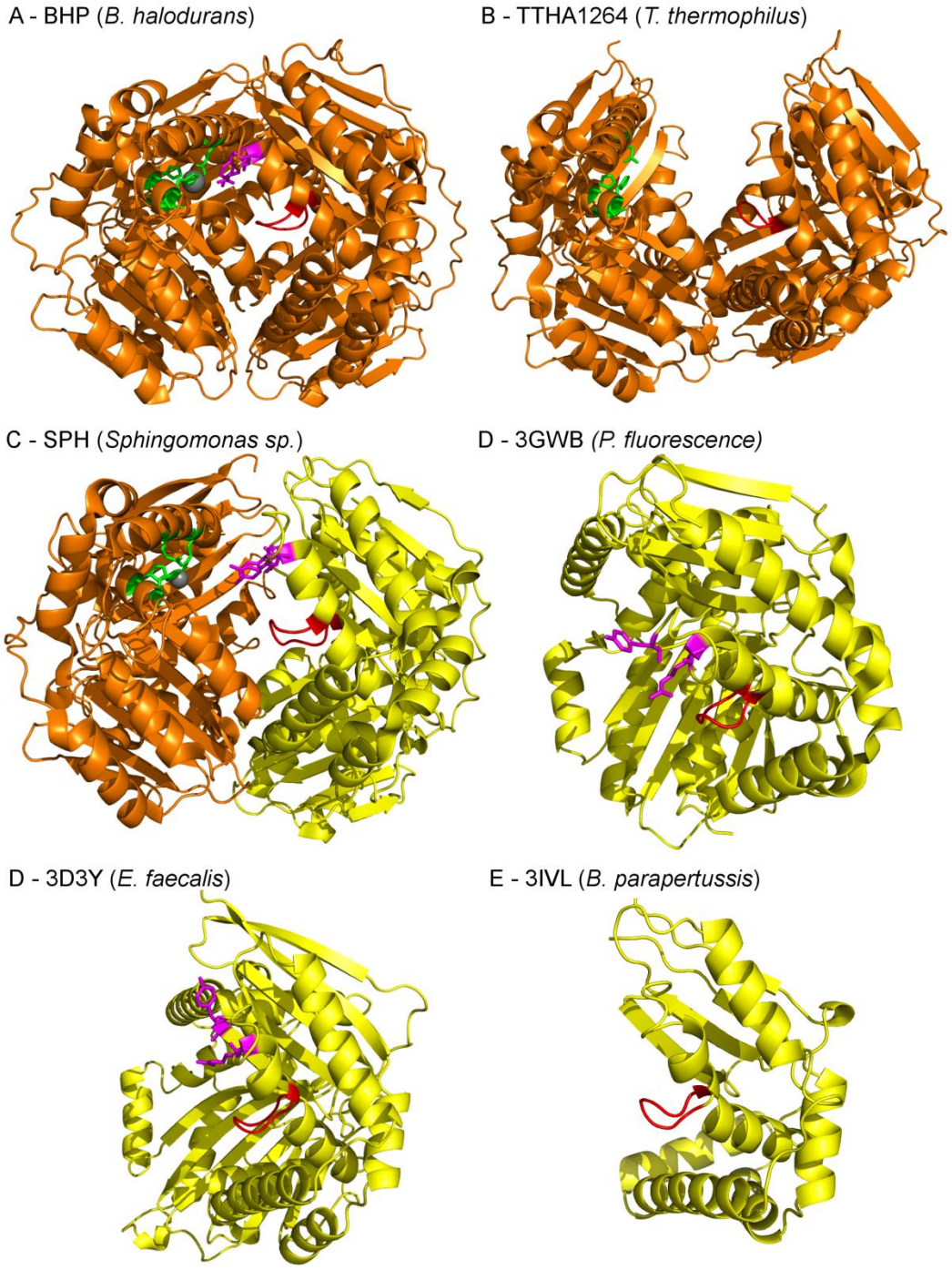


Figure 42. Crystal structures of selected M16 family bacterial peptidases. "Embryonal" form of GRL, RY motif and zinc-binding motif are displayed in red, magenta and green color, respectively. Zinc-binding subunits (β-MPP-like) are displayed in orange and RY motif or "embryonal GRL" bearing subunits (α-MPP-like) are displayed in yellow color. Further details are in the text.

α-MPP *S. cerevisiae*	GGGGSFSAGGPGKGMYSRL
α-hpp *T. vaginalis*	GGGSEFSSEGLGSGFSSLL
GPP *G. intestinalis*	PETVLAFPVQPGRETDAAL
RPP *R. prowazekii*	FGG----------MSSRL
BHP *B. halodurans*	LGGS----------MSSRL
BH2393 *B. halodurans*	GG----------FSHSKL
TTHA1265 *T. thermophilus*	SGG----------MSSRL
α-SPH *Sphingomonas sp.*	LGGGG---------FESRL
3GWB *P. fluorescence*	LGGGG---------FGTRL
3D3Y *E. faecalis*	GIFGG---------FPHSK

Figure 43. Alignments of the GRL regions of yeast α-MPP and MPP-like proteins from selected organisms. The GRL residues 285-300 of the α-MPP subunit are underlined.

Kitada et al. stated that MPP may originate from a monomeric α-proteobacterial peptidase, similar to the *Rickettsia prowazekii* processing peptidase (RPP) [129]. Providing RPP is the farthermost known ancestor of MPP and considering the homology of RPP with YmxG (M16 family protease from *B. subtilis*), RPP might play a key role in negative regulation of the protein expression through its protease activity [141]. The first step in MPP evolution could be gene duplication leading to the homodimeric peptidasome, functionally similar to the *Bacillus halodurans* peptidase (BHP) and other members of the M16B.016 group. In the next step, the loss of one of the zinc-binding motifs and initial subunit specialization led to the creation of heterodimeric peptidases with a single active site, similar to the TTHA, SPH and other peptidases from *B. halodurans*, *P. fluorescence*, *E. faecalis* and *B. parapertussis* (i.e. the members of M16B.014, M16B.UNB and M16B.UPB group). Although all these bacterial peptidases do not possess the "full length" GRL, a typical functional trait of MPP, they do contain, in addition to the RY motif, an "embryonal" form of GRL in the subunit that might be an ancestor of α-MPP subunit (**Figure 43**) and thus might represent the immediate ancestor of MPP. However, in all these cases the "embryonal" GRL of bacterial peptidases does not fulfill the function of "full length" GRL of MPP (i.e. the substrate recognition).

In the next stage, the pathways diverged, although a heterodimeric organization with a single active site was maintained. One branch led to the eukaryotic M16B proteases, which have a dimeric organization very similar to the prokaryotic enzymes, but prolongation of the GRL of the "embryonal" to the "full length" form fixed both subunits in a stable open-close conformation and created new functionality of processing peptidase. The loss of the RY pair was then inevitable because it was no longer required for catalysis. The second branch led to the M16A/C peptidasomes, in which the two genes fused and were connected by a long linker. The presence of a covalent linkage relaxed the structural and mechanistic restraints needed to maintain a stable interface in the open conformation.

We hypothesize that the α-proteobacterial progenitor of mitochondrion already possessed a heterodimeric peptidase, similar to the contemporary SPH peptidase found in *Sphingomonas sp*. Our hypothesis is in contrast with Kitada et al.'s idea who claim the direct progenitor of MPP was the monomeric peptidase, similar to the RPP of *R. prowazekii*, and that homo- and/or heterodimeric formation of MPP evolving from RPP is subsequent to the moment of mitochondrion acquisition.

Another clue to the evolution of MPP provides the situation in plants where MPP is attached to the bc_1 complex in the form of Core I and Core II proteins. Bacteria have a bc_1 complex that lacks both Core I and II proteins. The different location of MPP in various eukaryotic organisms may represent divergence from a single original evolutionary event [109]. The evolution of the MPP and Core subunits could have started with an ancestral prokaryotic protease located in the cytosol of bacteria. During endosymbiosis, the processing peptidase might have become attached to the membrane as it was advantageous for the function of the early MPP to be located close to the protein import sites. Other gene duplication led to the detachment of MPP from the bc_1 complex in yeast and mammals and thus allowed independent regulation of the respiration and

mitochondrial import. The extra Core I and Core II subunits of bc_1 complex could have become necessary for protection against proteolytic degradation and for assembly of the bc_1 complex [158]. Thus, the Core proteins in yeast and mammals would be evolutionary relics of the processing peptidase [109]. Although Core II (homolog of α-MPP subunit) possesses a large loop, the sequence similarity with GRL is very poor (**Figure 43**). Indeed, the loop has only structural importance and does not fulfill the function on GRL of MPP (i.e. substrate recognition) since MPP is already present in the mitochondrial matrix.

We conclude that the ancestral MPP in the form of SPH-like heterodimeric peptidase was attached to the inner mitochondrial membrane in the vicinity of the future bc_1 complex where the prolongation of GRL from the "embryonal" to the "full length" form occurred. Later, gene duplication led to segregation of MPP from the evolving bc_1 complex while original subunits were converted to Core I and Core II proteins. This event allowed independent regulation of the processing activity and the other newly established function – the respiration. Indeed, even some contemporary organisms have one MPP subunit incorporated in bc_1 complex while second subunit is present in the soluble form on the matrix side of the inner mitochondrial membrane (*N. crassa* [125, 126], *D. discoideum* [127], *B. emersonii* [128]).

Furthermore we conclude that the presence or absence of the "full length" GRL can be considered as the evolutionary marker of the physiological function of the given MPP-like peptidase. Its biological impact confirms further the fact that GRL of MPP is almost absolutely conserved among various eukaryotic organisms, while the conservation of the regulatory α-subunit as a whole is reasonably lower if compared with the catalytic β-subunit.

The effect of GRL on MPP structure and function has been already discussed. Just to sum it up, GRL (i) keeps both MPP subunit in partially open-close conformation, regardless the presence of substrate in MPP active site, (ii) plays an active role during substrate translocation to the active site, and foremost,

(iii) is the place of primary interaction with substrate, i.e. recognizes hundreds of signal presequences of nucleus-encoded mitochondrial proteins. Specifically, hydrophobic interactions between GRL and signal presequences as a substrates play a crucial role in the substrate recognition process by MPP. The same interactions apply during protein translocation across the mitochondrial outer membrane, where the signal presequence forms an amphipathic helix and its hydrophobic side is recognized by Tom20 receptor. The Tom20 receptor is found in fungi and mammals but not in bacteria and thus was established *de novo* during transition of the mitochondrial α-proteobacterial ancestor to the contemporary form of mitochondria.

We conclude that while Tom20 represents an example of the host-directed (i.e. "outsiders" model) evolution of protein translocation machinery, MPP is an illustrative example of the organelle-driven (i.e. "insiders" model) evolutionary pathway of the eukaryotic cell evolution. Although both systems participate in different steps of the translocation process and do not exhibit homology, they developed independently principally the same mechanism of signal presequence recognition based not on the sequence similarity but on the functional analogy where hydrophobic interactions play a crucial role.

However, the evolution of mitochondrion did not stop at the moment when mitochondrial protein translocation machinery, including Tom20 and MPP, was established. Some eukaryotic cells don't possess mitochondria but evolutionary related reduced versions of mitochondria called hydrogenosomes and mitosomes. Indeed, MPP-like peptidases were identified in hydrogenosomes of *T. vaginalis* (HPP) and mitosomes of *G. intestinalis* (GPP) which constitute the next steps in the evolution of MPP. Unfortunately, the crystal structures of HPP and GPP are not known and our results regarding the structural features of HPP suggest that HPP is closely related to MPP and thus may have the same physiological function of processing peptidase. Indeed, according to sequence alignment (**Figure 43**), HPP contains modified but "full length" GRL and process

also mitochondrial signal presequences [131]. On contrary, GPP does not possess any form of GRL and thus represents very likely highly reduced version of MPP recognizing highly reduced set of signal presequences. In other words, GPP represents an example of reductive evolution of MPP leading to a primitive system with a different signal presequence recognition mechanism if compared with MPP or HPP. Interestingly, no receptor such as Tom20 or other components of the translocation channel of the TOM complex have so far been identified for *T. vaginalis* [159] or *G. intestinalis* [132].

It appears that while the principal selective pressure for the evolution of processing peptidases is probably their ability to efficiently cleave off the signal presequences, the differences in the properties of the substrate presequences may reflect in turn the different mechanism of their recognition on the outer organelle.

7. CONCLUSIONS

7.1 The role of GRL in MPP structure

An all atomic, non-restrained molecular dynamics simulation in explicit water was used to study in detail the structural features of the highly conserved glycine-rich loop of the α-subunit of the yeast MPP and its importance for the tertiary and quaternary conformation of MPP. Wild-type and GRL-deleted MPP structures were studied using non-restrained MD simulations, both in the presence and the absence of a substrate in the peptidase active site. Targeted MD simulations were employed to study the mechanism of substrate translocation from the GRL to the active site.

In this part, our findings are as follows:

- We demonstrated that the natural conformational flexibility of the GRL is crucial for the substrate translocation process from outside the enzyme towards the MPP active site.
- We showed that the α-helical conformation of the substrate is important not only during its initial interaction with MPP (i.e. substrate recognition) but also later, at least during the first third of the substrate translocation trajectory.
- Further, we showed that the substrate remains in contact with the GRL during at least the first half of the translocation trajectory and that hydrophobic interactions play a major role.
- Finally, we concluded that the GRL acts as a precisely balanced structural element, holding the MPP subunits in a partially closed conformation regardless the presence or absence of a substrate in its active site.

7.2 Structural features of HPP

Since the crystal structure of HPP has not been yet solved, our understanding of HPP structure and how the structural features define peptidase specificity is very limited. Thus, the first step towards understanding the HPP structure was construction of the homology model. The model was validated by advanced techniques such as molecular dynamics simulations, biological small-angle X-ray scattering, hydrogen-deuterium exchange and cross-linking studies. In addition, an attempt for HPP crystallization was performed.

In this part, our contributions are as follows:

- Expression and purification process for WT and E56Q mutant form of HPP was optimized and the importance of glycerol and DDM was emphasized.
- Crystallization experiments were performed with WT HPP without peptide substrate and such an experimental arrangement did not lead to proper protein crystals. Crystallization experiments with mutant form of HPP with peptide substrate are being performed.
- Homology model of HPP was built based on the crystal structure of homologous MPP and validated by both molecular dynamics simulation and cross-linking studies.
- Cross-linking studies showed very limited validity of homology model of HPP. Although the validity of individual β-HPP subunit is low, is still acceptable. On contrary, the homology structure of α-HPP is not valid and the same applies for quaternary organization of the HPP dimer. One cross-link identified between substrate and α-HPP suggests that α-subunit participates in substrate recognition and/or processing, similarly to the homologous MPP. Another intramolecular cross-link suggests that β-HPP subunit is capable of forming "head-to-tail" homodimers.

- Bio-SAXS experiments were performed both on WT and E56Q form of HPP. The effect of active-site-bound peptide substrate on tertiary and quaternary conformation of HPP was not confirmed. However, a structural difference between WT and E56Q form of HPP was shown. Homology model of HPP fits well the *ab initio* calculated solution shapes and the "tail extension" seemed to be artificial.

- Hydrogen-deuterium exchange confirmed all the results obtained by bio-SAXS method. In addition, the structural difference between WT and E56Q mutant form was assigned to the difference in β-HPP subunit.

7.3 Evolution of the GRL of MPP

We worked out the evolutionary scenario for GRL of MPP in the context of MPP evolution from ancestral bacterial peptidase to its reduced forms found in contemporary hydrogenosomes and mitosomes.

- We hypothesize that the α-proteobacterial progenitor of mitochondrion already possessed a heterodimeric peptidase, similar to the contemporary SPH peptidase found in *Sphingomonas sp*.

- Following mitochondrion acquisition the ancestral MPP in the form of SPH-like heterodimeric peptidase was attached to the inner mitochondrial membrane where the prolongation of GRL from the "embryonal" to the "full length" form occurred.

- Further, we conclude that the presence or absence of the "full length" GRL can be considered as the evolutionary marker of the physiological function of the given MPP-like peptidase. In other words, the "full length" GRL denotes the peptidase as a signal presequence processing peptidase.

- While Tom20 represents an example of the host-directed (i.e. "outsiders" model) evolution of protein translocation machinery, MPP is an illustrative

example of the organelle-driven (i.e. "insiders" model) evolution of the eukaryotic cell.

- We conclude that while hydrogenosomal HPP is functional homolog of MPP, the mitosomal GPP represents an example of strong reductive evolution of MPP leading back to a primitive peptidase, similar to bacterial MPP-like peptidases. Indeed, according to sequence alignment (**Figure 43**), while HPP contains very likely modified "full length" GRL, GPP does not possess any form of GRL. Moreover, HPP processes the same set of substrates as MPP while GPP does not.

8. PUBLICATION LIST

8.1 Publications with IF

Kučera T., Otyepka M., Matušková A., Samad A., Kutejová E., Janata J. (2013) *A computational study of the glycine-rich loop of mitochondrial processing peptidase*. PLoS ONE 8(9): e74518. doi:10.1371/journal.pone.0074518. **IF**: 4.090

Kadlčík S., **Kučera T.**, Chalupská D., Gažák R., Koběrská M., Ulanová D., Kopecký J., Kutejová E., Najmanová L., Janata J. (2013) *Adaptation of L-proline adenylation omain to rare substrate 4-propyl-L-proline in evolution of lincosamide antibiotics*. PLoS ONE 8(12): e84902. doi:10.1371/journal.pone.0084902. **IF**: 4.090

Dvořáková-Holá K., Matušková A., Kubala M., Otyepka M., **Kučera T.**, Večeř J., Heřman P., Parhomenko N., Kutejová E., Janata J. (2010) *Glycine-rich loop of mitochondrial processing peptidase alpha-subunit is responsible for substrate recognition by a mechanism analogous to mitochondrial receptor Tom20*. *Journal of Molecular Biology*. **396**, 1197-1210. **IF**: 4.008

Šmíd O., Matušková A., Harris S., **Kučera T.**, Novotný M., Horváthová L., Hrdý I., Kutejová E., Hirt R., Eembley T., Janata J. and Tachezy J. (2008) *Reductive evolution of the mitochondrial processing peptidases of the unicellular parasites Trichomonas vaginalis and Giardia intestinalis*. *PLOS PATHOGENS*. **4. IF**: 9.125

8.2 Publications without IF

Kutejová E., **Kučera T.**, Matušková A., Janata J. (2012) Chapter 262. Mitochondrial processing peptidase, *Handbook of Proteolytic Enzymes*, 3rd Edition, Elsevier.

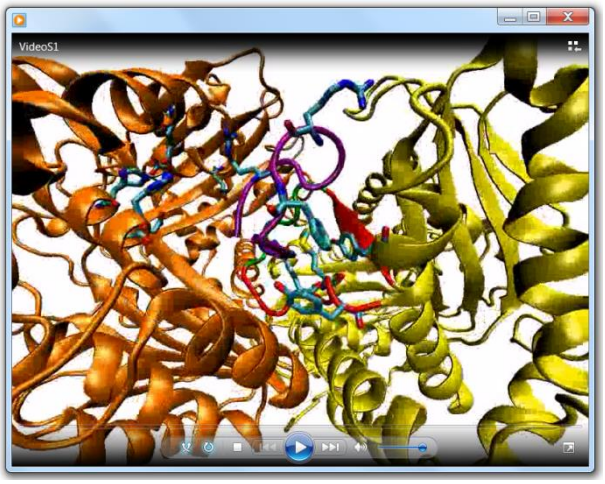
9. SUPPLEMNTARY MATERIAL

9.1 Multimedia

Package with multimedia data may be downloaded from:

- http://www.arecuk.cz/phd-thesis-multimedia.zip

9.1.1 Substrate translocation from GRL to MPP active site



Video 1. Targeted MD simulation of the substrate translocation from the place of its recognition (i.e. from the GRL) to the site of its processing (i.e. to the MPP active site). The restrain period is 1.6 ns and the total simulation time is 1.8 ns. α- and β-MPP subunits are displayed in yellow and orange color, respectively. The backbone trace of GRL is displayed as a red tube (residues 285-300) and K296, M298, Y299 and Y303 residues are shown in stick format. Peptide substrate is displayed as a tube in magenta color and R4, R8 (i.e. the R-2 motif) and F10 are shown in stick format. Zinc-binding residues H70, E73, H74 and E150 and the R-2-binding motif residues E160 and D164 of the β-MPP subunit are displayed as sticks. Zinc ion is shown as a blue sphere. Note that (i) the residues 289-393 of the GRL (green tube) are highly flexible and undergo major conformation changes during the whole substrate translocation process and that (ii) the R8 residue of the substrate is at the end of the simulation reoriented towards the E160 residue of the R-2-binding motif of the β-MPP subunit, just prior the subsequent proteolysis.

10. REFERENCES

1 Kimura, M. (1991) THE NEUTRAL THEORY OF MOLECULAR EVOLUTION - A REVIEW OF RECENT-EVIDENCE. Japanese Journal of Genetics. **66**, 367-386

2 Maher, K. A. and Stevenson, D. J. (1988) Impact frustration of the origin of life. Nature. **331**, 612-614

3 Cavalier-Smith, T. (1998) A revised six-kingdom system of life. Biological Reviews. **73**, 203-266

4 Woese, C. R. and Fox, G. E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. **74**, 5088-5090

5 Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. **87**, 4576-4579

6 Margulis, L. and Bermudes, D. (1985) Symbiosis as a mechanism of evolution: status of cell symbiosis theory. Symbiosis. **1**, 101-124

7 Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. and Woese, C. R. (1985) Mitochondrial origins. Proc Natl Acad Sci U S A. **82**, 4443-4447

8 Gray, M. W., Cedergren, R., Abel, Y. and Sankoff, D. (1989) On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci U S A. **86**, 2267-2271

9 Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H. and Kurland, C. G. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature. **396**, 133-140

10 Gray, M. W. (1998) Rickettsia, typhus and the mitochondrial connection. Nature. **396**, 109-110

11 Gray, M. W., Burger, G. and Lang, B. F. (1999) Mitochondrial evolution. Science. **283**, 1476-1481

12 Andersson, S. G. (1998) Bioenergetics of the obligate intracellular parasite Rickettsia prowazekii. Biochim Biophys Acta. **1365**, 105-111

13 Martin, W. and Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. Nature. **392**, 37-41

14 Andersson, S. G. and Kurland, C. G. (1999) Origins of mitochondria and hydrogenosomes. Curr Opin Microbiol. **2**, 535-541

15 Andersson, S. G. E., Karlberg, O., Canback, B. and Kurland, C. G. (2003) On the origin of mitochondria: a genomics perspective. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. **358**, 165-177

16 Vellai, T., Takacs, K. and Vida, G. (1998) A new aspect to the origin and evolution of eukaryotes. J Mol Evol. **46**, 499-507

17 Moreira, D. and Lopez-Garcia, P. (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol. **47**, 517-530

18 Lopez-Garcia, P. and Moreira, D. (1999) Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci. **24**, 88-93

19 Dyall, S. D. and Johnson, P. J. (2000) Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. Curr Opin Microbiol. **3**, 404-411

20 Boxma, B., de Graaf, R. M., van der Staay, G. W., van Alen, T. A., Ricard, G., Gabaldon, T., van Hoek, A. H., Moon-van der Staay, S. Y., Koopman, W. J., van Hellemond, J. J., Tielens, A. G., Friedrich, T., Veenhuis, M., Huynen, M. A. and Hackstein, J. H. (2005) An anaerobic mitochondrion that produces hydrogen. Nature. **434**, 74-79

21 Muller, M. (1993) The hydrogenosome. J Gen Microbiol. **139**, 2879-2889

22 Embley, T. M., van der Giezen, M., Horner, D. S., Dyal, P. L. and Foster, P. (2003) Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. **358**, 191-202

23 Embley TM, H. D., Hirt RP. (1997) Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? Trends Ecol Evol. **12**, 437-441

24 Dyall, S. D., Koehler, C. M., Delgadillo-Correa, M. G., Bradley, P. J., Plumper, E., Leuenberger, D., Turck, C. W. and Johnson, P. J. (2000) Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. Mol Cell Biol. **20**, 2488-2497

25 Bradley, P. J., Lahti, C. J., Plumper, E. and Johnson, P. J. (1997) Targeting and translocation of proteins into the hydrogenosome of the protist Trichomonas: similarities with mitochondrial protein import. Embo J. **16**, 3484-3493

26 Vellai, T. and Vida, G. (1999) The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. Proceedings of the Royal Society B-Biological Sciences. **266**, 1571-1577

27 Lang, B. F., Gray, M. W. and Burger, G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. Annu Rev Genet. **33**, 351-397

28 Biagini, G. A. and Bernard, C. (2000) Primitive anaerobic protozoa: a false concept? Microbiology. **146 (Pt 5)**, 1019-1020

29 Barton, R. M. and Worman, H. J. (1999) Prenylated prelamin A interacts with Narf, a novel nuclear protein. J Biol Chem. **274**, 30008-30018

30 Cavalier-Smith, T. (2006) Rooting the tree of life by transition analyses. Biol Direct. **1**, 19

31 Wong, J. T., Chen, J., Mat, W. K., Ng, S. K. and Xue, H. (2007) Polyphasic evidence delineating the root of life and roots of biological domains. Gene. **403**, 39-52

32 Forterre, P. (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol Microbiol. **33**, 457-465

33 Di Giulio, M. (2011) The last universal common ancestor (LUCA) and the ancestors of archaea and bacteria were progenotes. J Mol Evol. **72**, 119-126

34 Glansdorff, N., Xu, Y. and Labedan, B. (2008) The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct. **3**, 29

35 Embley, T. M. (2006) Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Philos Trans R Soc Lond B Biol Sci. **361**, 1055-1067

36 Cavalier-Smith, T. (1989) Molecular phylogeny. Archaebacteria and Archezoa. Nature. **339**, l00-01

37 Embley, T. M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. Nature. **440**, 623-630

38 Martijn, J. and Ettema, T. J. (2013) From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. Biochem Soc Trans. **41**, 451-457

39 Shimada, H. and Yamagishi, A. (2011) Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. Biochemistry. **50**, 4114-4120

40 Vesteg, M. and Krajčovič, J. (2011) The falsifiability of the models for the origin of eukaryotes. Curr Genet. **57**, 367-390

41 Woese, C. R. (2002) On the evolution of cells. Proc Natl Acad Sci U S A. **99**, 8742-8747

42 Guy, L. and Ettema, T. J. (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol. **19**, 580-587

43 Ettema, T. J., Lindås, A. C. and Bernander, R. (2011) An actin-based cytoskeleton in archaea. Mol Microbiol. **80**, 1052-1061

44 Yutin, N. and Koonin, E. V. (2012) Archaeal origin of tubulin. Biol Direct. **7**, 10

45 Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G. J., Hattori, M., Kanai, A., Atomi, H., Takai, K. and Takami, H. (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res. **39**, 3204-3223

46 Gabaldón, T. and Huynen, M. A. (2003) Reconstruction of the proto-mitochondrial metabolism. Science. **301**, 609

47 Gross, J. and Bhattacharya, D. (2011) Endosymbiont or host: who drove mitochondrial and plastid evolution? Biol Direct. **6**, 12

48 Gross, J. and Bhattacharya, D. (2009) Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective. Nat Rev Genet. **10**, 495-505

49 Alcock, F., Clements, A., Webb, C. and Lithgow, T. (2010) Evolution. Tinkering inside the organelle. Science. **327**, 649-650

50 Dolezal, P., Likic, V., Tachezy, J. and Lithgow, T. (2006) Evolution of the molecular machines for protein import into mitochondria. Science. **313**, 314-318

51 Hewitt, V., Alcock, F. and Lithgow, T. (2011) Minor modifications and major adaptations: the evolution of molecular machines driving mitochondrial protein import. Biochim Biophys Acta. **1808**, 947-954

52 Cavalier-Smith, T. (2006) Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. Proc Biol Sci. **273**, 1943-1952

53 Moran, N. A., McLaughlin, H. J. and Sorek, R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science. **323**, 379-382

54 Moran, N. A., McCutcheon, J. P. and Nakabachi, A. (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. **42**, 165-190

55 Karlberg, E. O. and Andersson, S. G. (2003) Mitochondrial gene history and mRNA localization: is there a correlation? Nat Rev Genet. **4**, 391-397

56 Herrmann, J. M. (2003) Converting bacteria to organelles: evolution of mitochondrial protein sorting. Trends Microbiol. **11**, 74-79

57 Kurland, C. G. and Andersson, S. G. (2000) Origin and evolution of the mitochondrial proteome. Microbiol Mol Biol Rev. **64**, 786-820

58 Gray, M. W., Burger, G. and Lang, B. F. (2001) The origin and early evolution of mitochondria. Genome Biol. **2**, REVIEWS1018

59 Lucattini, R., Likic, V. A. and Lithgow, T. (2004) Bacterial proteins predisposed for targeting to mitochondria. Mol Biol Evol. **21**, 652-658

60 Blobel, G. (1980) Intracellular protein topogenesis. Proc Natl Acad Sci U S A. **77**, 1496-1500

61 Moriwaki, K., Ogishima, T. and Ito, A. (1999) Analysis of recognition elements for mitochondrial processing peptidase using artificial amino acids: roles of the intervening portion and proximal arginine. J Biochem. **126**, 874-878

62 Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T. and Kohda, D. (2000) Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. Cell. **100**, 551-560

63 Saitoh, T., Igura, M., Obita, T., Ose, T., Kojima, R., Maenaka, K., Endo, T. and Kohda, D. (2007) Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. Embo J. **26**, 4777-4787

64 Burri, L., Williams, B. A., Bursac, D., Lithgow, T. and Keeling, P. J. (2006) Microsporidian mitosomes retain elements of the general mitochondrial targeting system. Proc Natl Acad Sci U S A. **103**, 15916-15920

65 Uboldi, A. D., Lueder, F. B., Walsh, P., Spurck, T., McFadden, G. I., Curtis, J., Likic, V. A., Perugini, M. A., Barson, M., Lithgow, T. and Handman, E. (2006) A mitochondrial protein affects cell morphology, mitochondrial segregation and virulence in Leishmania. Int J Parasitol. **36**, 1499-1514

66 Wickner, W. and Schekman, R. (2005) Protein translocation across biological membranes. Science. **310**, 1452-1456

67 Neupert, W. and Brunner, M. (2002) The protein import motor of mitochondria. Nat Rev Mol Cell Biol. **3**, 555-565

68 Rehling, P., Brandner, K. and Pfanner, N. (2004) Mitochondrial import and the twin-pore translocase. Nat Rev Mol Cell Biol. **5**, 519-530

69 Endo, T. and Kohda, D. (2002) Functions of outer membrane receptors in mitochondrial protein import. Biochim Biophys Acta. **1592**, 3-14

70 Meisinger, C., Ryan, M. T., Hill, K., Model, K., Lim, J. H., Sickmann, A., Müller, H., Meyer, H. E., Wagner, R. and Pfanner, N. (2001) Protein import channel of the outer mitochondrial membrane: a highly stable Tom40-Tom22 core structure differentially interacts with preproteins, small tom proteins, and import receptors. Mol Cell Biol. **21**, 2337-2348

71 Yamano, K., Yatsukawa, Y., Esaki, M., Hobbs, A. E., Jensen, R. E. and Endo, T. (2008) Tom20 and Tom22 share the common signal recognition pathway in mitochondrial protein import. J Biol Chem. **283**, 3799-3807

72 Moberg, P., Nilsson, S., Stahl, A., Eriksson, A. C., Glaser, E. and Maler, L. (2004) NMR solution structure of the mitochondrial F1beta presequence from Nicotiana plumbaginifolia. J Mol Biol. **336**, 1129-1140

73 Komiya, T., Rospert, S., Koehler, C., Looser, R., Schatz, G. and Mihara, K. (1998) Interaction of mitochondrial targeting signals with acidic receptor domains along the protein import pathway: evidence for the 'acid chain' hypothesis. EMBO J. **17**, 3886-3898

74 Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. and Pfanner, N. (2009) Importing mitochondrial proteins: machineries and mechanisms. Cell. **138**, 628-644

75 Neupert, W. and Herrmann, J. M. (2007) Translocation of proteins into mitochondria. Annu Rev Biochem. **76**, 723-749

76 Yi, L. and Dalbey, R. E. (2005) Oxa1/Alb3/YidC system for insertion of membrane proteins in mitochondria, chloroplasts and bacteria (review). Mol Membr Biol. **22**, 101-111

77 Herrmann, J. M. and Köhl, R. (2007) Catch me if you can! Oxidative protein trapping in the intermembrane space of mitochondria. J Cell Biol. **176**, 559-563

78 Schleiff, E. and Becker, T. (2011) Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat Rev Mol Cell Biol. **12**, 48-59

79 Koehler, C. M. (2004) New developments in mitochondrial assembly. Annu Rev Cell Dev Biol. **20**, 309-335

80 Rehling, P., Model, K., Brandner, K., Kovermann, P., Sickmann, A., Meyer, H. E., Kühlbrandt, W., Wagner, R., Truscott, K. N. and Pfanner, N. (2003) Protein insertion into the mitochondrial inner membrane by a twin-pore translocase. Science. **299**, 1747-1751

81 Meinecke, M., Wagner, R., Kovermann, P., Guiard, B., Mick, D. U., Hutu, D. P., Voos, W., Truscott, K. N., Chacinska, A., Pfanner, N. and Rehling, P. (2006) Tim50 maintains the permeability barrier of the mitochondrial inner membrane. Science. **312**, 1523-1526

82 Chacinska, A., Lind, M., Frazier, A. E., Dudek, J., Meisinger, C., Geissler, A., Sickmann, A., Meyer, H. E., Truscott, K. N., Guiard, B., Pfanner, N. and Rehling, P. (2005) Mitochondrial presequence translocase: switching between TOM tethering and motor recruitment involves Tim21 and Tim17. Cell. **120**, 817-829

83 Mokranjac, D. and Neupert, W. (2005) Protein import into mitochondria. Biochem Soc Trans. **33**, 1019-1023

84 Rawlings, N. D., Barrett, A. J. and Bateman, A. (2010) MEROPS: the peptidase database. Nucleic Acids Res. **38**, D227-233

85 Becker, A. B. and Roth, R. A. (1992) An unusual active site identified in a family of zinc metalloendopeptidases. Proc Natl Acad Sci U S A. **89**, 3835-3839

86 Kurochkin, I. V. (2001) Insulin-degrading enzyme: embarking on amyloid destruction. Trends Biochem Sci. **26**, 421-425

87 Miller, B. C., Eckman, E. A., Sambamurti, K., Dobbs, N., Chow, K. M., Eckman, C. B., Hersh, L. B. and Thiele, D. L. (2003) Amyloid-beta peptide levels in brain are inversely correlated with insulysin activity levels in vivo. Proc Natl Acad Sci U S A. **100**, 6221-6226

88 Taylor, A. B., Smith, B. S., Kitada, S., Kojima, K., Miyaura, H., Otwinowski, Z., Ito, A. and Deisenhofer, J. (2001) Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. Structure. **9**, 615-625

89 Oblong, J. E. and Lamppa, G. K. (1992) Identification of two structurally related proteins involved in proteolytic processing of precursors targeted to the chloroplast. EMBO J. **11**, 4401-4409

90 Stahl, A., Moberg, P., Ytterberg, J., Panfilov, O., Brockenhuus Von Lowenhielm, H., Nilsson, F. and Glaser, E. (2002) Isolation and identification of a novel mitochondrial metalloprotease (PreP) that degrades targeting presequences in plants. J Biol Chem. **277**, 41931-41939

91 Duckworth, W. C., Bennett, R. G. and Hamel, F. G. (1998) Insulin degradation: progress and potential. Endocr Rev. **19**, 608-624

92 Garcia, J. V., Gehm, B. D. and Rosner, M. R. (1989) An evolutionarily conserved enzyme degrades transforming growth factor-alpha as well as insulin. J Cell Biol. **109**, 1301-1307

93 Safavi, A., Miller, B. C., Cottam, L. and Hersh, L. B. (1996) Identification of gamma-endorphin-generating enzyme as insulin-degrading enzyme. Biochemistry. **35**, 14318-14325

94 Bennett, R. G., Duckworth, W. C. and Hamel, F. G. (2000) Degradation of amylin by insulin-degrading enzyme. J Biol Chem. **275**, 36621-36625

95 Qiu, W. Q., Walsh, D. M., Ye, Z., Vekrellis, K., Zhang, J., Podlisny, M. B., Rosner, M. R., Safavi, A., Hersh, L. B. and Selkoe, D. J. (1998) Insulin-degrading enzyme regulates extracellular levels of amyloid beta-protein by degradation. J Biol Chem. **273**, 32730-32738

96 von Heijne, G., Steppuhn, J. and Herrmann, R. G. (1989) Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem. **180**, 535-545

97 Miura, S., Amaya, Y. and Mori, M. (1986) A metalloprotease involved in the processing of mitochondrial precursor proteins. Biochem Biophys Res Commun. **134**, 1151-1159

98 Johnson, K. A., Bhushan, S., Ståhl, A., Hallberg, B. M., Frohn, A., Glaser, E. and Eneqvist, T. (2006) The closed structure of presequence protease PreP forms a unique 10,000 Angstroms3 chamber for proteolysis. EMBO J. **25**, 1977-1986

99 Glaser, E. and Alikhani, N. (2010) The organellar peptidasome, PreP: a journey from Arabidopsis to Alzheimer's disease. Biochim Biophys Acta. **1797**, 1076-1080

100 Aleshin, A. E., Gramatikova, S., Hura, G. L., Bobkov, A., Strongin, A. Y., Stec, B., Tainer, J. A., Liddington, R. C. and Smith, J. W. (2009) Crystal and solution structures of a prokaryotic M16B peptidase: an open and shut case. Structure. **17**, 1465-1475

101 Shen, Y., Joachimiak, A., Rosner, M. R. and Tang, W. J. (2006) Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. Nature. **443**, 870-874

102 Nagao, Y., Kitada, S., Kojima, K., Toh, H., Kuhara, S., Ogishima, T. and Ito, A. (2000) Glycine-rich region of mitochondrial processing peptidase alpha-subunit is essential for binding and cleavage of the precursor proteins. J Biol Chem. **275**, 34552-34556

103 Ohtsuka, J., Ichihara, Y., Ebihara, A., Nagata, K. and Tanokura, M. (2009) Crystal structure of TTHA1264, a putative M16-family zinc peptidase from Thermus thermophilus HB8 that is homologous to the beta subunit of mitochondrial processing peptidase. Proteins. **75**, 774-780

104 Alper, G. and Narayanan, V. (2003) Friedreich's ataxia. In Pediatr Neurol ed.)^eds.). pp. 335-341

105 Edland, S. D. (2004) Insulin-degrading enzyme, apolipoprotein E, and Alzheimer's disease. J Mol Neurosci. **23**, 213-217

106 Edland, S. D., Slager, S. and Farrer, M. (2004) Genetic association studies in Alzheimer's disease research: challenges and opportunities. Stat Med. **23**, 169-178

107 McCord, L. A., Liang, W. G. G., Dowdell, E., Kalas, V., Hoey, R. J., Koide, A., Koide, S. and Tang, W. J. (2013) Conformational states and recognition of amyloidogenic peptides of human insulin-degrading enzyme. Proceedings of the National Academy of Sciences of the United States of America. **110**, 13827-13832

108 Bohni, P., Gasser, S., Leaver, C. and Schatz, G. (1980) A matrixlocalized mitochondrial protease processing cytoplasmicallymade precursors to mitochondrial proteins. . ed.)^eds.). pp. 423–433, Elsevier

109 Braun, H. P. and Schmitz, U. K. (1995) Are the 'core' proteins of the mitochondrial bc1 complex evolutionary relics of a processing protease? Trends Biochem Sci. **20**, 171-175

110 Kitada, S., Shimokata, K., Niidome, T., Ogishima, T. and Ito, A. (1995) A putative metal-binding site in the beta subunit of rat mitochondrial processing peptidase is essential for its catalytic activity. J Biochem. **117**, 1148-1150

111 Striebel, H. M., Rysavy, P., Adamec, J., Spizek, J. and Kalousek, F. (1996) Mutational analysis of both subunits from rat mitochondrial processing peptidase. Arch Biochem Biophys. **335**, 211-218

112 Janata, J., Hola, K., Kubala, M., Gakh, O., Parkhomenko, N., Matuskova, A., Kutejova, E. and Amler, E. (2004) Substrate evokes translocation of both domains in the mitochondrial processing peptidase alpha-subunit during which the C-terminus acts as a stabilizing element. In Biochem Biophys Res Commun ed.)^eds.). pp. 211-217

113 Geli, V., Yang, M. J., Suda, K., Lustig, A. and Schatz, G. (1990) The MAS-encoded processing protease of yeast mitochondria. Overproduction and characterization of its two nonidentical subunits. J Biol Chem. **265**, 19216-19222

114 Luciano, P., Geoffroy, S., Brandt, A., Hernandez, J. F. and Geli, V. (1997) Functional cooperation of the mitochondrial processing peptidase subunits. J Mol Biol. **272**, 213-225

115 von Heijne, G. (1986) Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. **5**, 1335-1342

116 Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H. and Schatz, G. (1986) A chemically synthesized pre-sequence of an imported mitochondrial protein can form an amphiphilic helix and perturb natural and artificial phospholipid bilayers. EMBO J. **5**, 1327-1334

117 Komuro, Y., Miyashita, N., Mori, T., Muneyuki, E., Saitoh, T., Kohda, D. and Sugita, Y. (2013) Energetics of the presequence-binding poses in mitochondrial protein import through Tom20. J Phys Chem B. **117**, 2864-2871

118 DVORAKOVA-HOLA, K., MATUSKOVA, A., KUBALA, M., OTYEPKA, M., KUCERA, T., VECER, J., HERMAN, P., PARKHOMENKO, N., KUTEJOVA, E. and JANATA, J. (2010) Glycine-Rich Loop of Mitochondrial Processing Peptidase alpha-Subunit Is Responsible for Substrate Recognition by a Mechanism Analogous to Mitochondrial Receptor Tom20. Journal of Molecular Biology. **396**, 1197-1210

119 Nishino, T. G., Kitano, K., Kojima, K., Ogishima, T., Ito, A. and Kitada, S. (2007) Spatial orientation of mitochondrial processing peptidase and a preprotein revealed by fluorescence resonance energy transfer. J Biochem. **141**, 889-895

120 Amata, O., Marino, T., Russo, N. and Toscano, M. (2011) A proposal for mitochondrial processing peptidase catalytic mechanism. J Am Chem Soc. **133**, 17824-17831

121 Gakh, O., Cavadini, P. and Isaya, G. (2002) Mitochondrial processing peptidases. Biochim Biophys Acta. **1592**, 63-77

122 Deng, K., Zhang, L., Kachurin, A. M., Yu, L., Xia, D., Kim, H., Deisenhofer, J. and Yu, C. A. (1998) Activation of a matrix processing peptidase from the crystalline cytochrome bc1 complex of bovine heart mitochondria. J Biol Chem. **273**, 20752-20757

123 Deng, K., Shenoy, S. K., Tso, S. C., Yu, L. and Yu, C. A. (2001) Reconstitution of mitochondrial processing peptidase from the core proteins (subunits I and II) of bovine heart mitochondrial cytochrome bc(1) complex. J Biol Chem. **276**, 6499-6505

124 Xia, D., Yu, C. A., Kim, H., Xian, J. Z., Kachurin, A. M., Zhang, L., Yu, L. and Deisenhofer, J. (1997) Crystal structure of the cytochrome bc(1) complex from bovine heart mitochondria (vol 277, pg 60, 1997). Science. **278**, 2037-2037

125 Schulte, U., Arretz, M., Schneider, H., Tropschug, M., Wachter, E., Neupert, W. and Weiss, H. (1989) A family of mitochondrial proteins involved in bioenergetICS and biogenesis. Nature. **339**, 147-149

126 Arretz, M., Schneider, H., Guiard, B., Brunner, M. and Neupert, W. (1994) Characterization of the mitochondrial processing peptidase of Neurospora crassa. J Biol Chem. **269**, 4959-4967

127 Nagayama, K., Itono, S., Yoshida, T., Ishiguro, S., Ochiai, H. and Ohmachi, T. (2008) Antisense RNA inhibition of the beta subunit of the Dictyostelium discoideum mitochondrial processing peptidase induces the expression of mitochondrial proteins. Bioscience Biotechnology and Biochemistry. **72**, 1836-1846

128 Rocha, C. R. and Gomes, S. L. (1999) Characterization and submitochondrial localization of the alpha subunit of the mitochondrial processing peptidase from the aquatic fungus Blastocladiella emersonii. J Bacteriol. **181**, 4257-4265

129 Kitada, S., Uchiyama, T., Funatsu, T., Kitada, Y., Ogishima, T. and Ito, A. (2007) A protein from a parasitic microorganism, Rickettsia prowazekii, can cleave the signal sequences of proteins targeting mitochondria. J Bacteriol. **189**, 844-850

130 Brown, M. T., Goldstone, H. M., Bastida-Corcuera, F., Delgadillo-Correa, M. G., McArthur, A. G. and Johnson, P. J. (2007) A functionally divergent hydrogenosomal peptidase with protomitochondrial ancestry. Mol Microbiol. **64**, 1154-1163

131 SMID, O., MATUSKOVA, A., HARRIS, S., KUCERA, T., NOVOTNY, M., HORVATHOVA, L., HRDY, I., KUTEJOVA, E., HIRT, R., EMBLEY, T., JANATA, J. and TACHEZY, J. (2008) Reductive Evolution of the Mitochondrial Processing Peptidases of the Unicellular Parasites Trichomonas vaginalis and Giardia intestinalis. PLOS PATHOGENS. **4**

132 Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G. and Sogin, M. L. (2007) Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science. **317**, 1921-1926

133 Dolezal, P., Smid, O., Rada, P., Zubacova, Z., Bursac, D., Sutak, R., Nebesarova, J., Lithgow, T. and Tachezy, J. (2005) Giardia mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. Proc Natl Acad Sci U S A. **102**, 10924-10929

134 Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R. and Samuelson, J. (1999) Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite Entamoeba histolytica. Mol Cell Biol. **19**, 2198-2205

135 Tovar, J., Fischer, A. and Clark, C. G. (1999) The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica. Mol Microbiol. **32**, 1013-1021

136 Williams, B. A., Hirt, R. P., Lucocq, J. M. and Embley, T. M. (2002) A mitochondrial remnant in the microsporidian Trachipleistophora hominis. Nature. **418**, 865-869

137 Mach, J., Poliak, P., Matusková, A., Zárský, V., Janata, J., Lukes, J. and Tachezy, J. (2013) An Advanced System of the Mitochondrial Processing Peptidase and Core Protein Family in Trypanosoma brucei and Multiple Origins of the Core I Subunit in Eukaryotes. Genome Biol Evol. **5**, 860-875

138 Dabonné, S., Moallic, C., Sine, J. P., Niamké, S., Dion, M. and Colas, B. (2005) Cloning, expression and characterization of a 46.5-kDa metallopeptidase from Bacillus halodurans H4 sharing properties with the pitrilysin family. Biochim Biophys Acta. **1725**, 136-143

139 Maruyama, Y., Chuma, A., Mikami, B., Hashimoto, W. and Murata, K. (2011) Heterosubunit composition and crystal structures of a novel bacterial M16B metallopeptidase. J Mol Biol. **407**, 180-192

140 Chen, N. Y., Jiang, S. Q., Klein, D. A. and Paulus, H. (1993) ORGANIZATION AND NUCLEOTIDE-SEQUENCE OF THE BACILLUS-SUBTILIS DIAMINOPIMELATE OPERON, A CLUSTER OF GENES ENCODING THE 1ST 3 ENZYMES OF DIAMINOPIMELATE SYNTHESIS AND DIPICOLINATE SYNTHASE. Journal of Biological Chemistry. **268**, 9448-9465

141 Bolhuis, A., Koetje, E., Dubois, J. Y., Vehmaanperä, J., Venema, G., Bron, S. and van Dijl, J. M. (2000) Did the mitochondrial processing peptidase evolve from a eubacterial regulator of gene expression? Mol Biol Evol. **17**, 198-201

142 Yabuuchi, E., Yano, I., Oyaizu, H., Hashimoto, Y., Ezaki, T. and Yamamoto, H. (1990) PROPOSALS OF SPHINGOMONAS-PAUCIMOBILIS GEN-NOV AND COMB NOV, SPHINGOMONAS-PARAPAUCIMOBILIS SP-NOV, SPHINGOMONAS-YANOIKUYAE SP-NOV, SPHINGOMONAS-ADHAESIVA SP-NOV, SPHINGOMONAS-CAPSULATA COMB NOV, AND 2 GENOSPECIES OF THE GENUS SPHINGOMONAS. Microbiology and Immunology. **34**, 99-119

143 Kucera, T. (2008) Charakterizace signálních peptidas mitochondriálního typu ed.)^eds.)

144 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics. **23**, 2947-2948

145 Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. **31**, 3381-3385

146 Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. and Svergun, D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography. **40**, s223-s228

147 Svergun, D., Barberato, C. and Koch, M. H. J. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography. **28**, 768-773

148 Wriggers, W. (2010) Using Situs for the integration of multi-resolution structures. Biophys Rev. **2**, 21-27

149 Kavan, D. and Man, P. (2011) MSTools—Web based application for visualization and presentation of HXMS data. International Journal of Mass Spectrometry. **302**, 53-58

150 Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W. and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. Proceedings of the National Academy of Sciences. **97**, 5802-5806

151 Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. (2005) The Amber biomolecular simulation programs. J. Comput. Chem. **26**, 1668-1688

152 Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. **65**, 712-725

153 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. J Mol Graph. **14**, 33-38, 27-38

154 Popov, A. V., Vorobjev, Y. N. and Zharkov, D. O. (2013) MDTRA: a molecular dynamics trajectory analyzer with a graphical user interface. J Comput Chem. **34**, 319-325

155 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol. **372**, 774-797

156 The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC. ed.)^eds.)

157 Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol. **179**, 125-142

158 Boumans, H., Berden, J. A. and Grivell, L. A. (1997) The role of subunit VIII in the structural stability of the bc1 complex from Saccharomyces cerevisiae studied using hybrid complexes. Eur J Biochem. **249**, 762-769

159 Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Müller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M. and Johnson, P. J. (2007) Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science. **315**, 207-212